Department
for Environment
Food & Rural Affairs

Implementation of the Nitrate Pollution Prevention Regulations 2015 in England

Method for designating Nitrate Vulnerable Zones for surface freshwaters

December 2016

This publication is available at www.gov.uk/government/publications

Any enquiries regarding this publication should be sent to us at

enquiries@environment-agency.gov.uk

www.gov.uk/defra

Contents

1. Introduction and overview 1

1.1 Background 1

1.2 The review of surface water Nitrate Vulnerable Zones 1

1.3 Evolution of assessment methodology 2

2. Key Concepts 3

2.1 Continuous designations 3

2.2 Nitrogen chemistry in rivers 4

2.3 What form of nitrogenous compounds do we refer to in this assessment? 5

2.4 Why do we use the 95th percentile to characterise water quality for NVZs? 6

2.5 Types of surface water NVZ designation 7

2.6. What happens when the monitoring location is within the mixing zone of a significant point source? 9

2.7 The role of agriculture in NVZ designations 10

3. Overview of the surface water method 14

4. Monitoring Data 18

4.1 Statistical analysis of water quality monitoring data 18

4.2 Representativeness of monitoring points 27

5. Land Use Modelling 30

5.1 Data 31

5.2 Modelling method 35

5.3 Model checking 38

5.4 Model use 38

6. Combining Monitoring and Modelling results in the surface water evidence matrix 42

6.1 Lines of evidence 42

6.2 The evidence matrix 43

6.3 Converting evidence matrix scores into NVZ areas.....44

6.4 Case study: applying the evidence matrix to the Essex Stour catchment.....45

7. De-designation.....47

8. Using local knowledge to test draft designations.....52

8.1 Case study: Discussing a new designation where the monitoring failure is downstream of a significant point source.....54

9. Integrating areas of change into existing NVZ boundaries.....56

9.1 Rules for removing designations.....57

9.2 Modification of boundaries to better represent hydrology.....58

References.....62

Appendix A. WIMS data extraction criteria.....64

Appendix B. Water company monitoring data.....65

Appendix C. Multiple Outlier Test.....67

C1 Objective.....67

C2 Definitions.....67

C3 The single outlier test.....67

C4 The Multiple Outlier Test.....69

Appendix D: Note on the Weibull and Quantile regression methods.....70

D1 The Weibull method.....70

D2 Quantile regression.....70

1. Introduction and overview

1.1 Background

The Nitrates Directive (91/676/EEC) aims to protect water quality by preventing nitrogenous compounds from agricultural sources polluting groundwater and surface waters and by promoting the use of good farming practices. We are required under the terms of this Directive to identify waters which are or could become polluted by nitrogenous compounds and to designate these waters and all contributing land that drains to them as Nitrate Vulnerable Zones (NVZs). Farmers in designated areas must follow an Action Programme to reduce nitrogen pollution from agricultural sources.

This document describes the method applied to determine the status of surface waters during the most recent NVZ review. This is referred to as the 2017 review because the resulting designations are due to come into force in January 2017.

1.2 The review of surface water Nitrate Vulnerable Zones

The Nitrates Directive is concerned with the level of pollution due to nitrogenous compounds in surface waters and groundwater, and whether coastal or freshwaters are eutrophic. Article 3 of the Nitrates Directive and the relevant Annex set the following criteria for identifying polluted surface waters:

- *whether surface freshwaters, in particular those used or intended for the abstraction of drinking water, contain or could contain, if action pursuant to Article 5 is not taken, more than the concentration of nitrates laid down in accordance with EU Drinking Water legislation (75/440/EEC)*

Polluted, in the context of the Drinking Water Directive (98/83/EC) that defines the threshold value above which human health is at risk, means that the 95th percentile nitrogen concentration in the water is more than 50 mg/l as nitrate. This is the threshold that we use to assess surface water quality under the Nitrates Directive.

The Nitrates Directive requires us to review NVZ designations at least every four years. The role of each review is two-fold:

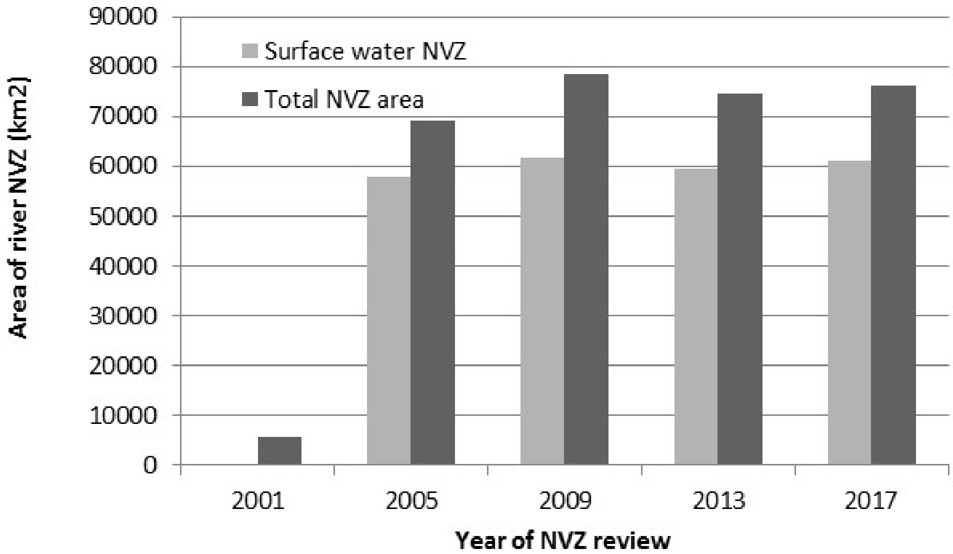
1. To review the evidence supporting existing NVZ designations

2. To identify any areas where water quality has deteriorated to such an extent that these waters are, or could be, affected by pollution.

The periodic nature of reviewing NVZs means that each review necessarily presents a 'snapshot' of nitrate pollution up to the time of the review. This latest review uses data up to and including 2014.

1.3 Evolution of assessment methodology

This is the sixth review of NVZs in England. Surface Water designations have remained relatively stable since the 2009 review, with a few areas being removed and a few others added. Surface water NVZ designations however remain the single largest component of the overall area designated as NVZ in England. The following graph shows the changes in designated surface water area over time and the contribution to the total NVZ area[1].



•

Figure 1 - Area of surface water NVZ over time

The surface water NVZ method has evolved over time to take advantage of new analytical techniques and data. Here is a summary of the most significant changes to the surface water NVZ method.

The 2005 review was based only on monitoring data and designated all land upstream of polluted monitoring sites as NVZ. It took no account of the land use within those areas.

The 2009 review built on the 2005 method by considering how land use influences losses of nitrogen. The 2009 method development was assisted by an expert panel, which also included stakeholders such as the NFU – the Method Review Group (MRG). Method developments agreed by this group included:

[1] The first review prior to 2001 simply designates NVZ areas and does not distinguish between groundwater and surface water designations. It is therefore not possible to include figures on the area of surface water NVZ only from the first NVZ assessment.

- Comparing monitoring data with modelled sources of nitrogen using an evidence matrix

- Using WFD river catchments as the units of designation

- Using confidence intervals to classify the risk of waters being polluted rather than simply considering them to either pass or fail.

The 2009 method was re-examined by the MRG ahead of the 2013 review to ensure it included the latest scientific understanding and modelling techniques. The most significant changes being:

- statistical analysis of the monitoring data incorporated quantile regression techniques for increased robustness

- use of Internal Drainage Districts as well as WFD river catchments as units of assessment in areas draining to the Wash

For the 2017 review, further refinements were made to the method and these are highlighted in this document. Since any changes are refinements and not large changes to the 2013 method, the MRG was not asked to review the method for 2017.

2. Key Concepts

The following key concepts are integral to understanding the 2017 review method.

2.1 Continuous designations

NVZ reviews are not undertaken in isolation and NVZ designations are principally a continuation of the outcome of the last review, unless the evidence shows that change is needed.

Every four years we analyse all available data on nitrogen levels in rivers but if an area has been designated previously**, it does not need to meet the designation criteria in the current review for the designation to be retained**. This is because the Nitrates Directive requires us to designate land draining to waters that are affected by pollution *or could be affected in the future*. If an area has been designated it means that there is or has been a risk of nitrogen pollution. Unless certain criteria are met then that risk remains. We discuss de-designation in more detail in Section 7.

The majority of NVZ designations are those that were designated previously and where water quality has not changed enough to trigger de-designation. In many cases water quality has, however, shown signs of improvement and we acknowledge this in the evidence we provide to support each designation.

2.2 Nitrogen chemistry in rivers

Nitrogen is present in different forms in freshwaters and the relationship between these different forms is dynamic. This means that the proportion of the different forms changes depending on the oxygen content of the water or river bed and the presence of different bacteria. Article 2 in the Nitrates Directive defines the forms of nitrogen which should be measured with the statement *'nitrogen compound: means any nitrogen-containing substance except for gaseous molecular nitrogen'*

The most common nitrogen compounds in freshwaters are nitrate (NO_3), nitrite (NO_2) and ammonium (NH_4^+). Ammonium is produced from the breakdown of urea (and other nitrogenous wastes) in sewage treatment works (STWs); from livestock; from nitrogenous fertilisers applied to land; or it may be applied directly as a fertiliser. It can also form when ammonia gas (often released from large animal units) is reduced on contact with the soil.

Ammonium is present in all water bodies as part of the total nitrogen load. In the presence of oxygen it is rapidly transformed to nitrite and then to nitrate. This change is called nitrification. Nitrification plays a crucial role in increasing the movement of nitrogen in the environment (Ayre et al, 1997). It is a two-stage, aerobic process involving bacteria such as *Nitrosomonas* and *Nitrobacter*. In rivers, the majority of the nitrification occurs on the stream bed.

The first stage is the conversion of ammonium to nitrite by *Nitrosomonas* spp bacteria:

$$NH_4 + 1.5O_2 \Leftrightarrow NO_2^- + H_2O + 2H^+$$

The second stage is the conversion of nitrite to nitrate by *Nitrobacter* spp

$$NO_2^- + \frac{1}{2}O_2 \Leftrightarrow NO_3^-$$

Both of these conversions are rapid and the process is dynamic. The reverse reaction, called denitrification, involves the reduction of nitrate to produce nitrogen gas and is, in contrast, slow. In low oxygen conditions excess nitrate may also be converted back to ammonia (ammonification).

Denitrification generally occurs where oxygen is depleted, and bacteria respire nitrate as a substitute. Due to the high concentration of oxygen in our atmosphere, denitrification only takes place where oxygen consumption exceeds supply and where sufficient quantities of nitrate are present. Such conditions may occur naturally in anoxic or extremely low oxygen conditions in fine, organic-rich, lowland river bed sediments.

Denitrification happens more quickly in the low-oxygen soil environment when compared to rivers, especially where the rivers are fast flowing. Ponding and wetland construction are frequently used to provide a suitable environment for de-nitrification in point source treatment systems. Denitrification generally proceeds through some combination of the following intermediate forms of nitrogen compounds:

$NO_3^- \rightarrow NO_2^- \rightarrow NO + N_2O \rightarrow N_2$ (g)

The complete denitrification process can be expressed as a redox reaction:

$2\ NO_3^- + 10\ e^- + 12\ H^+ \rightarrow N_2 + 6\ H_2O$

We use rate coefficients to estimate the rate of change of substances into something else (e.g. ammonia into nitrate). This is often called 'decay' and these coefficients are unitless. A decay rate of 0.1 per day means a tenth of the substance changes into something else every day. This means that the rate of change takes into account different starting conditions, only removing a proportion of what is already there. This means that, as in natural systems like rivers, the rate of 'decay' of a substance slows as the amount of the substance present decreases.

Rates for nitrification are given in a range of standard texts (USEPA rates, constants and kinetics, Chapra 1997), academic papers and model manuals (Eatherall *et al* 1998, Wade 2006). Ranges are given in most texts, with rates generally between 1 and 9 per day. Rates vary along the river length and by season; depending on the amount of ammonia initially present; the suitability of the bed substrate for bacteria; the temperature; and dissolved oxygen levels. Conversion of ammonia to nitrate is fastest in shallow streams with pebbly beds which provide a large surface area for bacterial growth.

In contrast, the rate of de-nitrification is much slower, generally given as 0.1 per day although some authors estimate rates lower than this, eg Wade 2006 (0.04 to 0.09 per day).

In summary, ammonium is very rapidly converted to nitrate in freshwater systems, and nitrate is slowly removed as nitrogen gas.

2.3 What form of nitrogenous compounds do we refer to in this assessment?

The Nitrates Directive uses the terms nitrates and compounds of nitrogen. This can cause confusion as the standard is given as 50 mg/l of **nitrate**. When we measure nitrogen compounds, including nitrate, we measure the compound as the nitrogen content in the sample, i.e. mgN/l. This is because for water quality monitoring purposes, we are interested in all the nitrogen compounds, and we sum these per sample to calculate losses in terms of **Total Inorganic Nitrogen (TIN).** When we look at our analysis of the monitoring data in chapter 4, you will see that all the determinands we analyse are the various nitrogen compounds as N, e.g. nitrate as N (expressed as NO_3 - N).

Nitrogen, and therefore nitrate as N, is much lighter than nitrate because of the additional oxygen in the nitrate compound. To relate the values of TIN we use in our assessment to the Nitrates Directive standard of 50mg/l of nitrate, we must convert TIN to NO_3 using the atomic weight of nitrogen (14) compared to the total weight of a nitrate molecule (62). This means that nitrogen is 22.6% of the total weight of the nitrate molecule and;

- To calculate the nitrogen content of a measurement of nitrate, you must multiply by 14/62 (22.6%)
- To calculate the weight of the nitrate molecule with a nitrogen content of x mg/l, you must multiply this number by 4.43 ($\frac{1}{(14/62)}$).

So 50 mg/l as NO_3 = 11.3 mg/l as N or TIN

On all graphs and charts that follow, the value of 11.3 mg/l as TIN is the level that is used to show that water is polluted.

2.4 Why do we use the 95th percentile to characterise water quality for NVZs?

A percentile is a summary statistic that provides information about the distribution (spread) of values in a defined population; for example, the sample data over time from a particular monitoring location. If you measured 100 values from a population, the 95%ile would be the value that was exceeded only 5% of the time. EC drinking water legislation stipulates a 95th percentile statistic. The 95th percentile is well-suited to standards where we need to be precautionary (where exceedence would risk harm to human health).

For the surface water method we need to make sure that levels of nitrogenous compounds in rivers do not exceed the standard of 11.3 mg/l as TIN or 50mg/l as NO_3. Figure 2.1 gives an example of a monitoring site where the Nitrates Directive standard is regularly breached, based on the 95th percentile. The data periods for the NVZ review rounds are shown as green lines. This site was first designated in 2004, when 13 out of 72 samples were above the standard (18% of samples). Between 2004 and 2009, 7 out of 61 samples were above the standard (11% of samples). And finally, between 2009 and 2014, 11 out of 56 samples were above the standard (20% of samples). So on each occasion, the monitoring site was above the standard based on the 95th percentiles of the review periods.

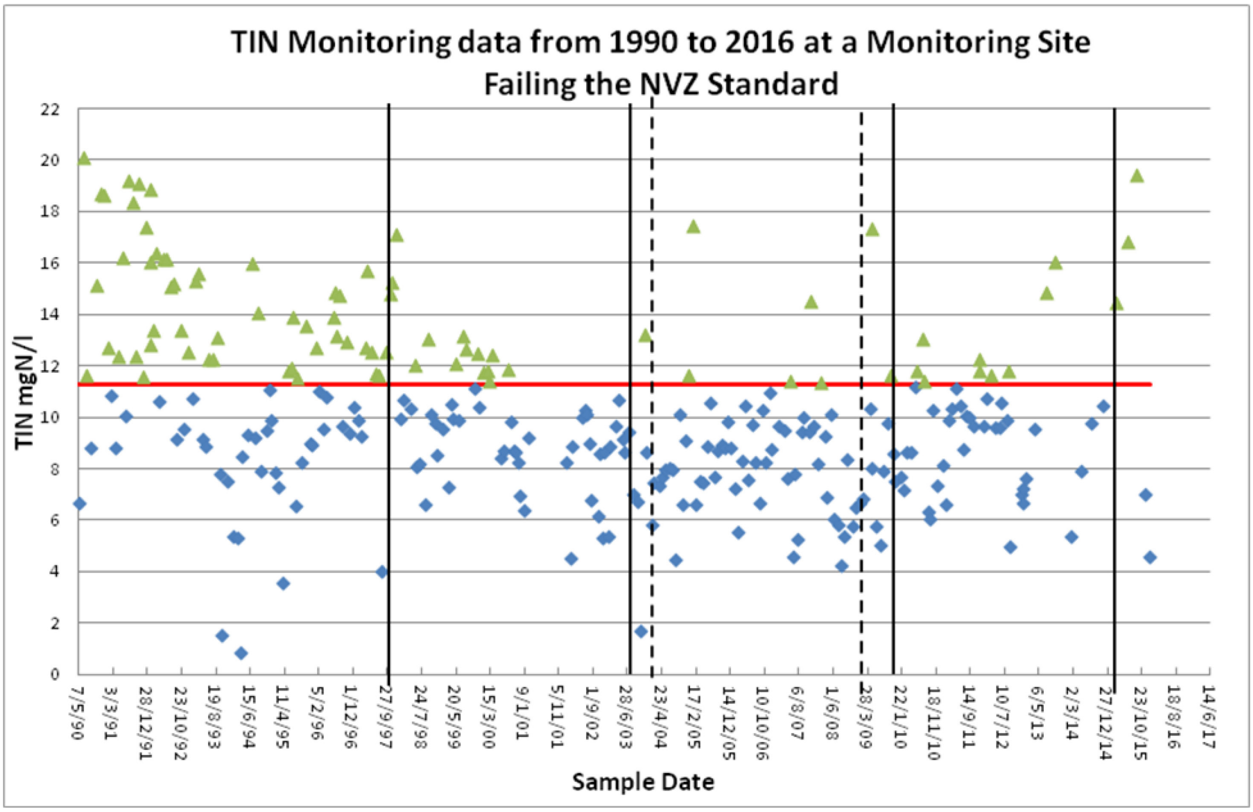


Figure 2.1 - Data from a monitoring site which fails the Nitrates Directive standard. Samples above 11.3 are shown in green, those below in blue. The NVZ designation cycles are shown as black lines (note that the years 2004 and 2009 overlap two designation periods)

2.5 Types of surface water NVZ designation

Since the 2009 review the unit of assessment for the surface water NVZ method has been the WFD river catchment. This splits the land area into units that drain to distinct parts of the river network (referred to as water bodies).

The river network in each WFD river catchment can be split into two categories:

1. The main river which receives inputs from all upstream parts of the catchment

2. Tributaries which drain to the main river but only receive inputs from the immediate local area.

Our monitoring network consists of sites on main rivers and tributaries, and the distinction is important when we define NVZs.

If the main river is considered to fail or to be at risk of failing, according to the NVZ method, then **all** land draining to the furthest downstream point (the outlet) within the WFD river catchment is designated a NVZ. We call this a **catchment designation.**

If a tributary is considered to fail or is at risk of failing then only the WFD river catchment in which it is located is designated. We call this a **water body designation.**

In both catchment and water body designations, the NVZ starts at the outlet of the failing WFD river catchment, regardless of how far up the WFD river catchment the monitoring point is located.

The following diagram illustrates this:

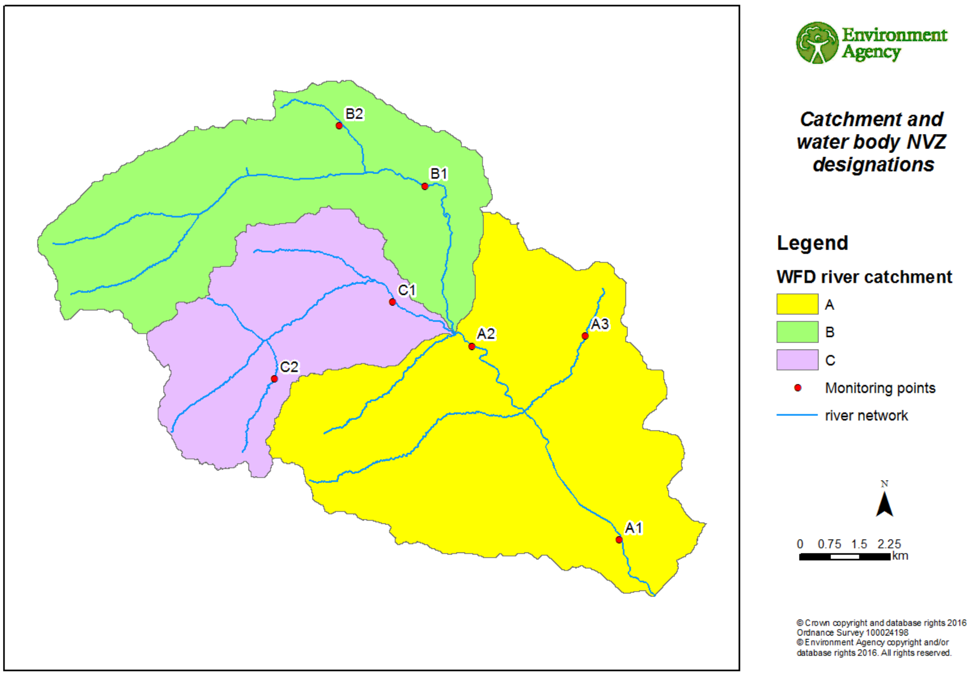


Figure 2.2 - Example WFD river catchments and monitoring points

In Figure 2.2 we see all the land draining to the outlet of a river. The area is split into three WFD river catchments: A, B and C. There are a number of monitoring points, some on the main river for each WFD river catchment (A1, A2, B1, C1) and some on tributaries (A3, B2 and C2). If any of these fail then the nature of the NVZ depends on which of the monitoring points fail (Table 2.1).

Table 2.1 - Type of surface water NVZ designations resulting from failures at monitoring points shown in Figure 2.2

Failing monitoring point	Resulting NVZ type	WFD river catchments designated
A1	Catchment NVZ	A, B and C
A2	Catchment NVZ	A, B and C
A3	Water body NVZ	A only
B1	Catchment NVZ	B only
B2	Water body NVZ	B only
C1	Catchment NVZ	C only
C2	Water body NVZ	C only

Whichever points in WFD river catchments B and C fail, the NVZ is the same size but a different type of designation. Also, even though A2 is a long way upstream of A1, it makes no difference to the size of the catchment NVZ because the WFD river catchment is our unit of assessment.

2.6. What happens when the monitoring location is within the mixing zone of a significant point source?

Monitoring locations within the mixing zone of a discharge with a numeric consent for nitrogen (either as ammonia or nitrate) are excluded from our assessment as the sample results may not be representative of the overall quality of the water body.

To exclude any monitoring location that is within a mixing zone, locations are screened according to EU guidance on setting mixing zones under the EQS Directive (*2008/105/EC*). Any monitoring locations which have been screened out in previous NVZ reviews are not re-used in the 2017 review.

The 2013 method mentions a distance of 1km downstream of a major point source as the trigger value for calculating the likely mixing zone length. This does not mean that 1km is an appropriate value for mixing zones downstream of all discharges.

In shallow running freshwaters, discharge mixing zones are generally quite short in length. The EU guidance (EU 2010) recommends that a mixing zone of 10 times the river width is a conservative estimate of the actual zone. For larger rivers, where a mixing zone is more complex, the industry standard CORMIX model (Doneker, R.L. and G.H. Jirka 2007) is generally used to calculate the mixing zone of a discharge.

The Environment Agency’s detailed river network gives the width of all recognised watercourses in England, and field measurements can be used when these are available for comparison. The following example of a mixing zone experiment illustrates the length of river required for complete mixing. In this experiment red dye was introduced into the river to simulate a pollution incident. In the first photograph (Figure 2.3.1) upstream from a bridge, you can see the dye entering a small river (less than 5m wide), and that it is not completely mixed with the river water. A monitoring location here would be inappropriate, as a sample might contain up to 100% effluent or 100% upstream river water, depending on where it was taken from.

Figure 2.3.2 below shows the river 50m downstream of the bridge, where the dye has completely mixed with the river water. Samples taken in this location will be representative of all upstream sources.


Figures 2.3.1 and 2.3.2 - Illustrations of a mixing zone in a small river

2.7 The role of agriculture in NVZ designations

Representative sample points

When land is within a NVZ, farms are subject to the NVZ Action Programme. Designations on the other hand are based on observed monitoring data or a statistical model that seeks to characterise the main sources of TIN within catchments. Both the monitoring and the modelling therefore include the contribution from agricultural sources but the decision to designate or not is based on TIN from all sources, not just agricultural ones.

The Nitrates Directive itself refers only to NVZs being required where the water is polluted, as long as the monitoring data is representative of the nitrogen pollution in the catchment.

It therefore aims to ensure that agriculture contributes to the improvement in areas draining to polluted waters but does not demand that agriculture is the only or dominant source in the catchment.

The Directive states in Article 6:

For the purpose of designating and revising the designation of vulnerable zones, Member States shall:

(a) within two years of notification of the Directive, monitor the nitrate concentration in freshwaters over a period of one year:

(i) at surface water sampling stations, laid down in Article 5 (4) of Directive 75/440/EEC and/or at other sampling stations which are representative of surface waters of Member States ....

As part of each review, we examine the locations of monitoring points and remove any judged not to be representative of the land draining to them (see Sections 2.6 and 4.5).

The contribution of agriculture to the total N load

Article 3 of the Directive requires member states to report:

1. Waters affected by pollution and waters which could be affected by pollution if action pursuant Article 5 is not taken shall be identified by the Member States in accordance with the criteria set out in Annex I.

2. Member States shall, within a two-year period following the notification of this Directive, designate as vulnerable zones all known areas of land in their territories which drain into the waters identified according to paragraph 1 and which contribute to pollution. They shall notify the Commission of this initial designation within six months.

The Directive itself does not quantify a minimum contribution which agriculture must make to the TIN level in the watercourse, only that the water is polluted and action to reduce the agricultural contribution can be taken. The current UK interpretation of the Directive, the Nitrate Pollution Prevention Regulations 2015, similarly does not quantify a minimum contribution from agriculture required for NVZ designation, only that pollution from agriculture and other sources be considered.

However, the NVZ designations are subject to an appeals process and in the appeals the issue of what is a 'significant contribution' has been raised. As neither the Directive nor the Regulations makes any reference to contributions from a particular sector, only that pollution from all sectors be considered, when designations are appealed we look to judgements of the European Courts for guidance.

The ruling of the ECJ in a case concerning the significance of the contribution of the Walloon area of agriculture to nitrogen pollution in the North Sea is frequently used as a reference when judging whether the agricultural contribution to the TIN in a catchment is 'significant' or not. (***EC v Belgium***, C-221/03, Judgment of the Court, 22 September 2005).

This judgement stated that '*Walloon agriculture contributes 19% of the total nitrogen in the Meuse basin and 17% of the total nitrogen in the Escaut basin. Those two rivers cross the Walloon Region and drain into the North Sea. It must be pointed out that, although minor, these contributions are by no means insignificant'*

While the ECJ judgement against Belgium does not address the issues of contributions from individual sectors, it does give us a guideline figure for what can be considered not insignificant. We consider this ruling consistent with the screening process we apply to individual monitoring points to ensure that the data used to designate NVZs are not unduly influenced by a single source. This is discussed in more detail in section 4.5.

- As part of the 2017 review, we present estimates of annual source apportionment in the supporting evidence for each NVZ. These are not core data for the designation method. They are however a useful additional test to check that the NVZs are justified in accordance with our responsibility under the Nitrates Directive. The source apportionment figures in section 5.1 and Appendix 2 of the datasheets are for guidance only and are based on national datasets. These may be refined using more detailed local datasets.

The example below illustrates a designation where there is a 'significant' contribution from both agriculture and a point source.

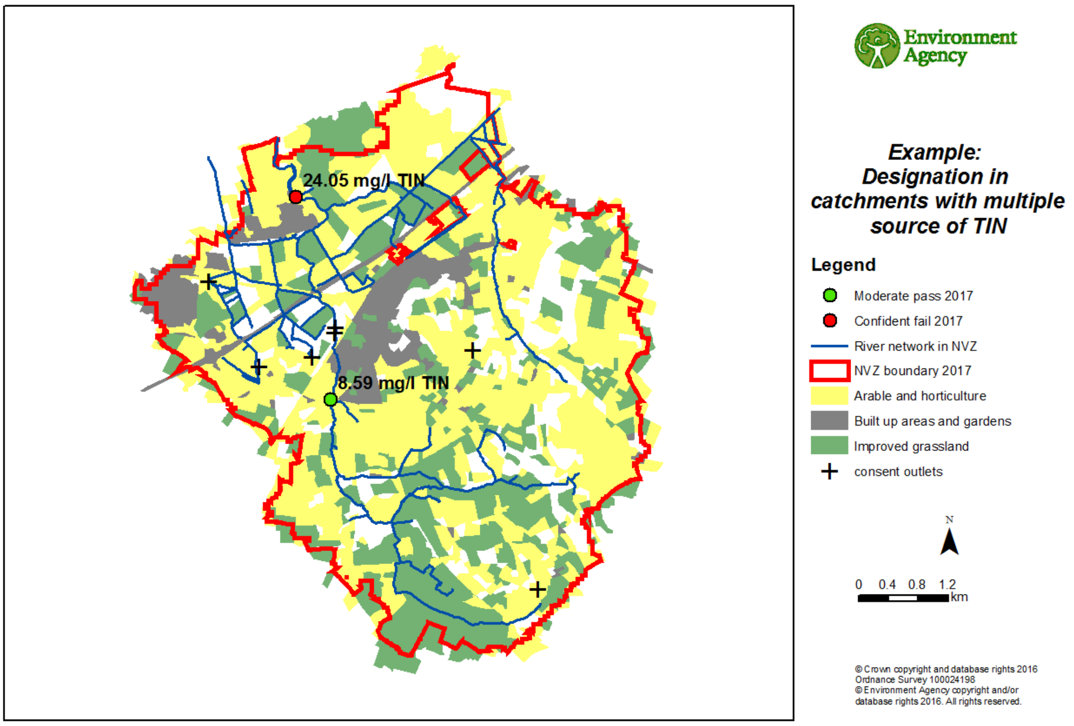


Figure 2.4 - Example NVZ where TIN levels result from both agriculture and other sources

In this catchment TIN levels above the STW are less than 11.3mg/l but above that level below the discharge. It has been argued in appeals that this means that the problem in this area is related to sewage treatment and a NVZ is not justified.

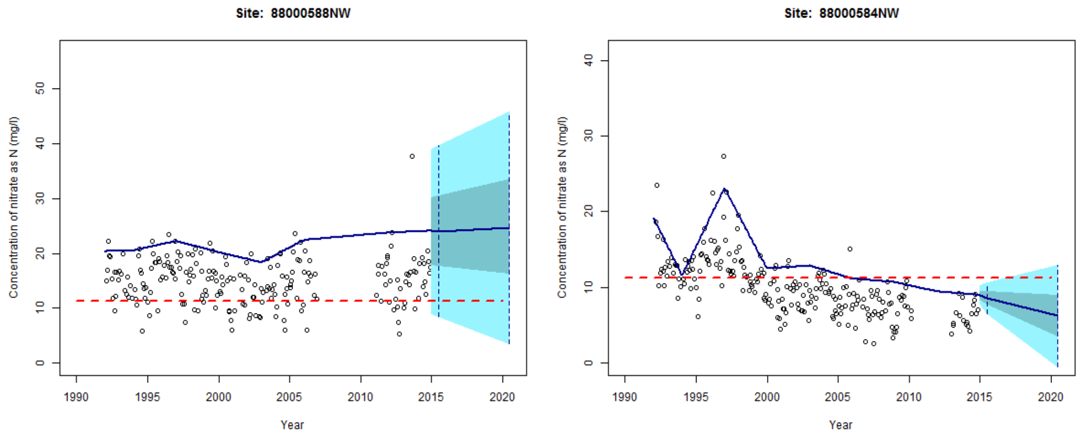
Whilst it is true that without the STW this catchment would not fail, it is incorrect to consider the STW in isolation. At the sample point above the STW the 95th percentile is 8.6mgN/l, which equates to 76% (8.6/11.3) of the TIN standard. Although this is not enough to be classified polluted by itself, it shows there is a high background concentration arising from the land above the STW.

The yellow and green areas are fields classified as either arable or permanent grassland according to the 2007 CEH land cover map of Great Britain. These are the two agricultural land uses that result in higher TIN losses so it is clear that agriculture makes a contribution to background TIN levels.

Whilst there are other consented discharges contributing to the upstream TIN (marked as crosses in Figure 2.4) the upstream land is predominantly agricultural. Agriculture and other sources therefore contribute to the elevated TIN in this catchment even though it is the discharge from the sewage works that causes levels to exceed 11.3mgN/l. The monitoring time series corroborates this picture and Figures 2.5.1 and 2.5.2 below show the monitoring data first below the sewage works and then from the monitoring point upstream. The monitoring downstream (Figure 2.5.1) shows that TIN is consistently higher than 11.3mgN/l and frequently exceeds 20 mgN/l. At the upstream monitoring point

(Figure 2.5.2), there are no recent values above 11.3 mgN/l but even the lowest values are 5 mgN/l or higher.

This indicates that regardless of conditions and which sources are contributing most at particular times, there is a significant background TIN concentration in the river.



Figures 2.5.1 and 2.5.2 - Monitoring data downstream and upstream of STW in our example NVZ

Therefore designation is required in this catchment in accordance with our responsibility to the Nitrates Directive.

3. Overview of the surface water method

The current surface water method can be separated into a number of distinct phases:

1. reviewing the current data, which includes both monitoring and modelled data
2. combining monitoring and modelled data to identify which waters are polluted or at risk of pollution according to our evidence matrix
3. reviewing these results to see:
    - where water quality has deteriorated beyond our thresholds for designation
    - where water quality has shown consistent and prolonged evidence of being at a level we can confidently say is ‘low risk’
    - where water quality is stable, or has deteriorated/improved but not to a sufficient level to be considered for the changes described in (a) or (b)

The designation process is divided into these three steps. The rest of this document describes in greater technical detail how we carry out each of the steps. Section 2 has introduced some ‘Key Concepts’ which aid the reader in understanding the methods and calculations used in the method. The diagram below, Figure 3.1, describes the process we follow in reviewing the NVZs every four years, and we use this as a map for the rest of this document.

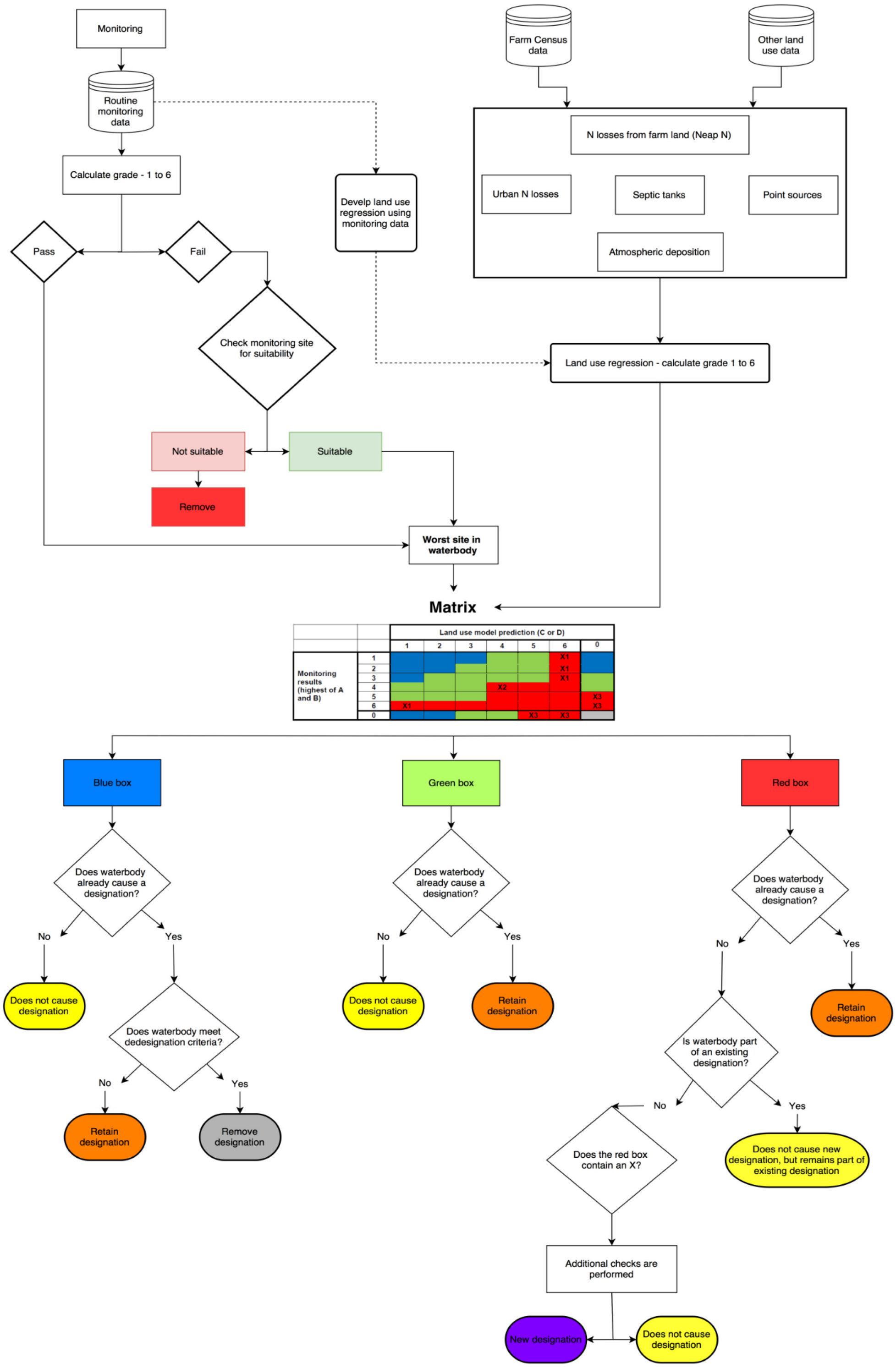


Figure 3.1 - Conceptual model of surface water NVZ method

The top half of the diagram describes the review of the data, where we investigate contemporary monitoring data and identify how many sites fail to meet the required standards of the Nitrates Directive. Any failing sites are subject to further scrutiny to ensure that they can be considered appropriate for use in the review of NVZs. Assuming that the failing sites are considered reliable indicators of water quality in the catchment then each WFD river catchment is characterised by the worst (i.e. highest TIN) monitoring site in the catchment.

Section 4 describes the analysis and screening of the monitoring data in more detail.

Monitoring data informs and is supported by land use modelling. This involves relating the monitoring data statistics to a range of data that describe the main sources of TIN in WFD river catchments. The land use modelling determines the best relationship between the catchment data and the monitoring data and predicts what TIN levels should be according to this relationship. Land use modelling serves two purposes:

1. it allows us to predict TIN levels where we lack monitoring data

2. it allows us to identify areas where the monitoring data is significantly lower or higher than we would expect based on the land use within that catchment

The land use modelling is described in more detail in Section 5.

At the centre of Figure 3.1 is the evidence matrix, in which we combine monitoring and land use modelling classes for each WFD river catchment and depending on the combination of those classes, determine if the NVZ status of the area should change or not. This is described in Section 6.

The lower half of Figure 3.1 describes the decisions that result from the review of current data.

Depending on the combination of monitoring and modelling classes, each WFD river catchment will be assigned to one of three categories:

1. Evidence supports the requirement for a new designation. These are shown as red boxes in the evidence matrix.

2. Evidence does not strongly indicate that any change of designation status is required. These are shown as green boxes.

3. Evidence indicates that water quality is at a level that we consider low risk. These are shown as blue boxes.

At this last stage we carry out further checks, especially when the evidence indicates that a change in NVZ status may be required. These further checks include discussion with local Environment Agency staff in a series of workshops. This part of the process is described in Section 8.

Having identified the areas where we are confident of a need for changed NVZ status, the last stage is to reconcile the NVZs from the current review with the previous review. This is described in Section 9.

4. Monitoring Data

Monitoring data is the principal type of evidence in the surface water method; it is used to assess the current state of water quality, predict what water quality will be in the future and it is an integral input to the land use modelling we describe in Section 5.

4.1 Statistical analysis of water quality monitoring data

We analyse water quality data to calculate a current and forecast estimate of the 95th percentile concentration of TIN. Both 95th percentile concentrations and the confidence intervals (CIs) around these concentrations are calculated for each sample point with sufficient data.

Based on these statistics we assign a current and future class to each point. How we assign these classes is illustrated in Figure 4.1 below.

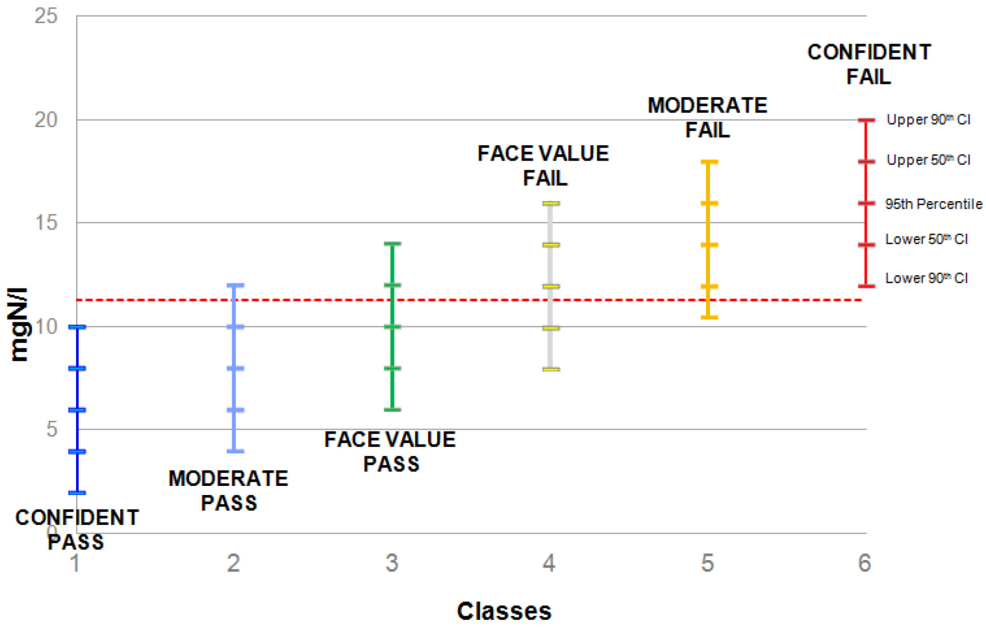


Figure 4.1 - NVZ monitoring classes based on the 95th percentile and confidence intervals

In this diagram we see how the 95th percentile estimates and their confidence intervals relate to the TIN standard (11.3mgN/l), shown here as the dashed red line. As more of the confidence intervals exceed the TIN standard, then the more confident we are that the monitoring point fails. The following table describes what each of these classes mean in terms of how confident we are that the monitoring point will fail the Nitrates Directive standards.

Table 4.1 - Monitoring classes

Class	**Meaning**
1	We are 95% confident that the 95th percentile is below 11.3
2	We are 75% confident that the 95th percentile is below 11.3
3	We have low (≥50%) confidence that the 95th percentile is below 11.3. The 95th percentile does not exceed 11.3, the lower confidence intervals are less than 11.3 but the upper confidence intervals are greater than 11.3
4	We have low (≥50%) confidence that the 95th percentile is

Class	Meaning
	above 11.3. The 95th percentile exceeds 11.3, the lower confidence intervals are less than 11.3 but the upper confidence intervals are greater than 11.3
5	We are 75% confident that the 95th percentile is above 11.3
6	We are 95% confident that the 95th percentile is above 11.3

The rest of this section gives an overview of the data used in the analysis, the analysis methods and the quality assurance performed on the results.

Data

We use data from the Environment Agency's Water Information Management System (WIMS) database and data provided by water supply companies from the period between 1st January 1990 and 31st December 2014.

Environment Agency data from all freshwater river sample points (which have not been removed from the analysis in previous rounds) with monitoring of appropriate determinands (substances) are used in the analysis. The determinands of interest are:

- ammoniacal nitrogen as N (WIMS determinand code 0111)
- total oxidised nitrogen as N (0116)
- nitrate as N (0117)
- nitrite as N (0118)

Results from monitoring locations for the same determinands from the same date range were requested from all English water supply companies via the industry body, Water UK. These data are added to the Environment Agency data to form a single merged dataset.

For more details on the extraction criteria, please see Appendix A and B.

Data processing

The merged dataset is subject to a data cleaning process to ensure all samples are suitable for use. We perform the following checks on the data:

Samples with zero values are removed. Zero values can occur for a number of reasons and it cannot be assumed that they represent readings below the Limit of Detection (LoD).

Samples with negative values are removed - a negative reading for the determinands listed above is always an error.

Outliers in the data are removed using the Multiple Outlier Test (MOT). See Appendix C for details.

Some monitoring points are equipped with auto-samplers which automatically collect water samples when defined criteria are met (e.g. a specified flow threshold). This leads to multiple samples with similar values which skews the dataset. To remove this bias from the dataset, we remove all samples collected by auto-samplers (identified as having three or more samples collected on the same day).

Where two samples were collected on the same day, we retain only the higher result.

Less than values in the data are treated using the standard Environment Agency approach of dividing the recorded value by two (Environment Agency Document 111_07_SD02 Water quality planning: Codes of practice for data handling).

To create the final dataset for analysis, the four determinands are combined to calculate TIN on each sampling occasion for each monitoring point. A final check of the number of TIN readings per monitoring point is performed. We exclude any monitoring points that have:

- fewer than 19 samples in total
- data from less than 5 calendar years
- no samples for the period 2009-2014

Statistical analysis

We analyse each monitoring point with sufficient data to determine whether or not:

- the current (2009 to 2014 or 2015) 95th percentile TIN concentration exceeds 11.3 mgN/l TIN as N (referred to as current TIN) or
- the future (2020) 95th percentile TIN concentration is likely to exceed 11.3 mgN/l TIN as N (referred to as future TIN)

If either the current or future TIN exceeded 11.3 mgN/l, the monitoring site is considered to have failed the assessment. The level of confidence in the result is recorded as one of the six classes shown in Table 4.1 above.

The statistical methods used to assess current and future TIN at each monitoring site depend on the quantity of data available, as set out in Table 4.2.

Table 4.2 - Data quantity rules for calculating current and future TIN

Rule	Current TIN assessment method	Future TIN assessment method
At least 19 samples and at least 5 years of data for period 2009-2014	*Weibull estimate* of 95th percentile TIN as N for 2009-2014	*Quantile regression* forecast of 95th percentile TIN as N in 2020
At least 19 samples and at least 5 years of data for period 1990-2014 and at least 1 sample for period 2009-2014	*Quantile regression* forecast of 95th percentile TIN as N in 2015	*Quantile regression* forecast of 95th percentile TIN as N in 2020
Fewer than 19 samples, fewer than 5 years of data for period 1990-2014, or no samples for period 2009-2014	No assessment	No assessment

Weibull method

The Weibull method is an established statistical technique for ranking data and calculating robust percentile estimates. See appendix D for a full description of this method and the Quantile Regression method described later. We use the Weibull method because it's relatively insensitive to outliers and doesn't require data to fit a particular distribution. The choice of a six year period (2009 to 2014) provides a good balance of responsiveness to change and lack of sensitivity to short term fluctuations.

An example of how the Weibull estimates relate to the data is shown below in Figure 4.2, where:

- The dark green solid horizontal line 2009-2014 marks the Weibull 95th percentile estimate
- The green shaded band indicates the 50% confidence interval around the Weibull 95th percentile estimate and
- The wider, light green shaded band indicates the 90% confidence interval around the Weibull 95th percentile estimate

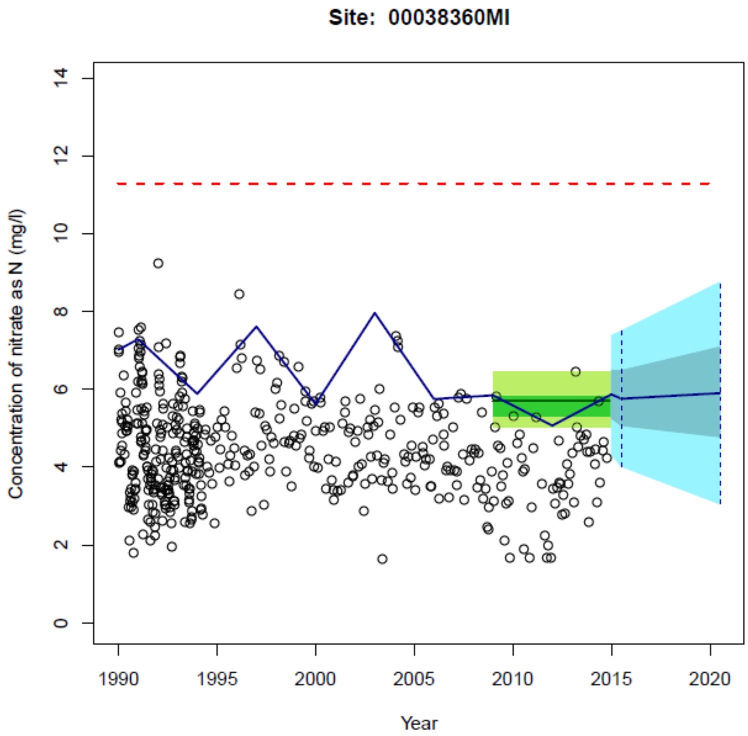


Figure 4.2 - Current 95th percentile prediction using the Weibull method

Quantile regression method

For all sites that met the minimum data requirements (at least 19 samples and at least 5 years of data for the period 1990-2014 and at least 1 sample for the period 2009-2014) we use quantile regression to characterise trends in water quality between 1990 and 2014. We then extrapolate the historical trend using Monte Carlo simulation[2] to estimate the current (2015) 95th percentile TIN as N concentration at monitoring points that have too few samples for the Weibull method, and to forecast the 95th percentile concentration in 2020.

Trend analysis

We fit a quantile regression model to the historical data (1990-2014) to describe variation in the 95th percentile TIN as N concentration at each monitoring site over time. The quantile regression uses a series of short-term trends, which are called linear splines[3], to predict the long term trend. Each spline is a minimum of three years or 12 samples long, whichever is longer. Each spline starts and ends at the end of a calendar year and at least three years from the beginning or end of the time series. The choice of a minimum time span of 3 years gives a balance between responsiveness and stability.

[2] Monte Carlo simulation is the process of repeating calculations using random sampling from a distribution to represent the uncertainty in the system.
[3] A spline is a curve that connects two or more points. In our case the spline is constrained to be linear, so is a linear slope that connects two points.

All the linear spline terms are carried through into the forecasting phase (Section 4.1.5.2), regardless of whether or not they were statistically significant. This is because they still provide the best estimates of the historical trends and if they are not statistically significant the effect on the outcome of the forecasting is likely to be small.

Forecasting

We assume that the 25 years (1990 to 2014) of historical data provide the best evidence as to the likely future direction of water quality and so we use Monte Carlo simulation to extrapolate the historical trend and forecast future water quality at each site. The forecasting model uses several parameters derived from the quantile regression for each monitoring point.

Suppose a quantile regression model (4.1.5.1) has *m* splines. In general, some of these slopes will have been estimated less precisely than others, which we can tell by their standard errors (SE). The forecasting simulation allows for this by using weights based on the SE of each spline which reduces the risk that our forecast is unduly influenced by short term changes in water quality.

The forecasting model is defined in the equation below;

Forecasting model

$$B_{i+1} = (1 - R_1) \times AvB + R_1 \times B_i + \sqrt{(1 - R_1{}^2)} \times (B_{rand} - AvB)$$

The model has the following terms;

- Normal distribution N(AvB, SDB) defined by weighted mean (AvB) and weighted standard deviation (SDB) of the m spline slopes. Weights are calculated for each j spline slope to be 1/SE(j)
- The final spline slope
- The lag-1 autocorrelation[4] of the spline slopes, R1, which defaults to 0 if there are fewer than five splines
- The average spline length (the mean number of years spanned per spline)
- The quantile regression prediction of the 95th percentile at the end of the monitored period (τ) and its standard error (τSE)

[4] Autocorrelation is correlation between points in a series. Here we are talking about one sample being correlated with a previous, earlier sample.

Taking as its starting point a 95th percentile concentration drawn from the Normal distribution N(τ, $τ_{SE}$) and the final spline slope, the Monte Carlo simulation generates a projection of how the 95th percentile TIN as N concentration might change in the future, starting from the year immediately following the end of the monitoring record.

The future projection is constructed by selecting a series of spline slopes at random (B_{rand}) from the Normal distribution N(AvB, SDB) and loosely associating each with the preceding slope to an extent governed by the autocorrelation coefficient R_1. Specifically, the slope for time step i+1 is generated from the previous time step's slope (B_i) and the random slope (B_{rand}) as in equation 1.

Taking the average spline length as the time step, the process is continued until the required forecasting horizon (2020) is reached. We repeat the whole process 10,000 times (all parameters are static bar B_{rand}) to generate a range of possible 'futures'. A variety of summary statistics are then calculated from each of the 10,000 simulations, including the median, which represents the best estimate of the 95th percentile concentration in 2015 and 2020. The uncertainty in this estimate is quantified by ranking the 10,000 forecasts for that year (2015 or 2020) and determining the 10th, 25th, 75th and 90th percentiles.

Figure 4.3 shows an example of a fitted quantile regression model and the forecasts generated from it. The red dashed line at 11.3 mgN/l indicates the threshold value and the solid blue zig-zag line indicates historical fluctuations in the 95th percentile TIN as N concentration. Beyond the end of the monitoring record in 2014:

- The blue line indicates the 95th percentile TIN as N concentration forecast by Monte Carlo simulation
- The blue shaded band indicates the 50% confidence interval around the 95th percentile TIN as N estimate
- The wider, light blue shaded band indicates the 90% confidence interval around the 95th percentile TIN as N estimate

In this example, the 95th percentile TIN as N concentration is forecast to be 5.8 mg N/l in 2015 (the first vertical dashed blue line) with a 90% confidence interval of 4.0-7.5 mg N/l, and 5.9 mgN/l in 2020 (the second vertical dashed blue line) with a 90% confidence interval of 3.1-8.8 mgN/l.

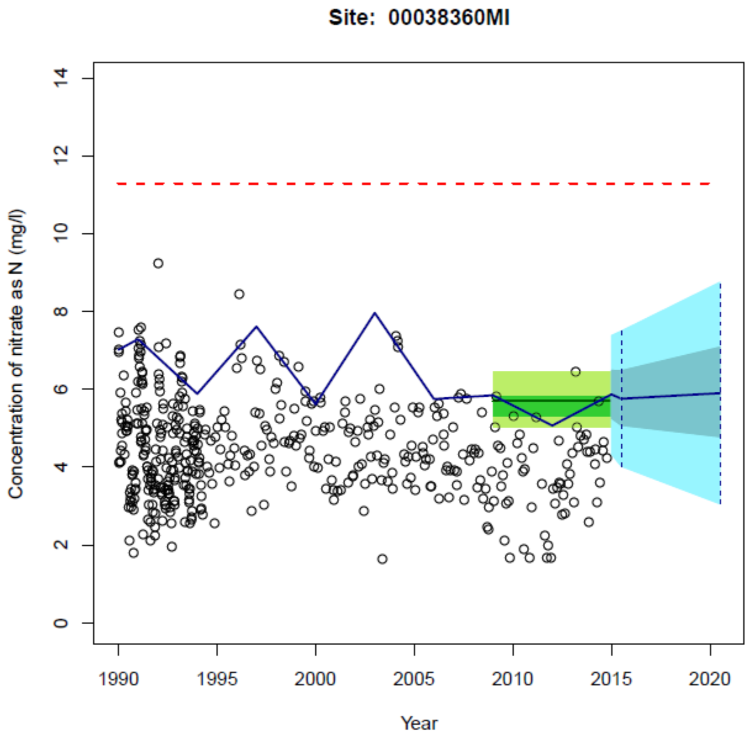


Figure 4.3 - Historical trend and future forecast of 95th percentile TIN concentration using quantile regression, and comparison to the 11.3 mgN/l TIN as N standard (red dashed line)

Quality assurance of statistical method

We undertake a series of checks to assess the performance of the statistical methods and the accuracy of current and trend TIN assessments. Monitoring points are flagged for more detailed manual checks if they display possible data quality issues, including;

- unusually high average or maximum concentrations
- unusually high coefficient of variation
- statistically significant bimodal or multi-modal distributions. These are rare in surface water quality data and may indicate that water quality is unduly affected by specific conditions at certain times of year. Our estimates of the 95th percentile will not reflect this so we must investigate these points further.
- large gaps in the monitoring record, or no recent monitoring data
- unusually wide confidence intervals around the results
- fail current or future status tests despite no samples exceeding the 11.3 mgN/l threshold

- pass the current and future status tests but have exceedances of the 11.3 mgN/l threshold since 2009

- large discrepancy between the Weibull and quantile regression estimates of the current 95^{th} percentile concentration

- unusually high or low quantile regression estimates of current or future TIN

In general, our checks confirmed that the statistical methodology has been correctly applied and that the current and trend TIN assessments are reasonable at the vast majority of monitoring points. Any monitoring points that fail to meet these criteria and would result in new designations are marked for discussion at local workshops (see Section 8).

4.2 Representativeness of monitoring points

We perform a series of further checks[5] on sample points with a class of 4 or higher (i.e. current or future TIN as N 95^{th} percentile exceeds 11.3 mgN/l) to determine that they are representative of the catchment they monitor: These checks include;

1. Is the monitoring location within the mixing zone of a point source discharge? (see key concept 2.6)

2. Are the 95^{th} percentile TIN as N estimates consistent with monitoring points up and downstream, and with monitoring points with similar land uses?

3. Are the observed 95^{th} percentile TIN as N estimates consistent with expectations given the catchment of the monitoring location?

4. Is the catchment of the monitoring point wholly urban?

5. Does ammonia form a significant portion of total TIN as N concentration at the monitoring location?

If a monitoring point is found to be within the mixing zone of a point source discharge, it is removed from the designation dataset. If we are not satisfied after performing checks 2 -5 that the monitoring point is representative of its catchment, then the monitoring location is either removed from the designation dataset or discussed in the local workshops (see Section 8). Only check 1 automatically results in the exclusion of the monitoring point from our analysis. The remaining checks are suggestive but not conclusive.

[5] In addition to the quality assurance of the results detailed in 4.1

Figure 4.4 summarises the process for checking monitoring points. It is not exhaustive but we undertake further checks on the validity of a designation at later stages either when we apply the evidence matrix (see Section 6) or during local workshops (see Section 8).

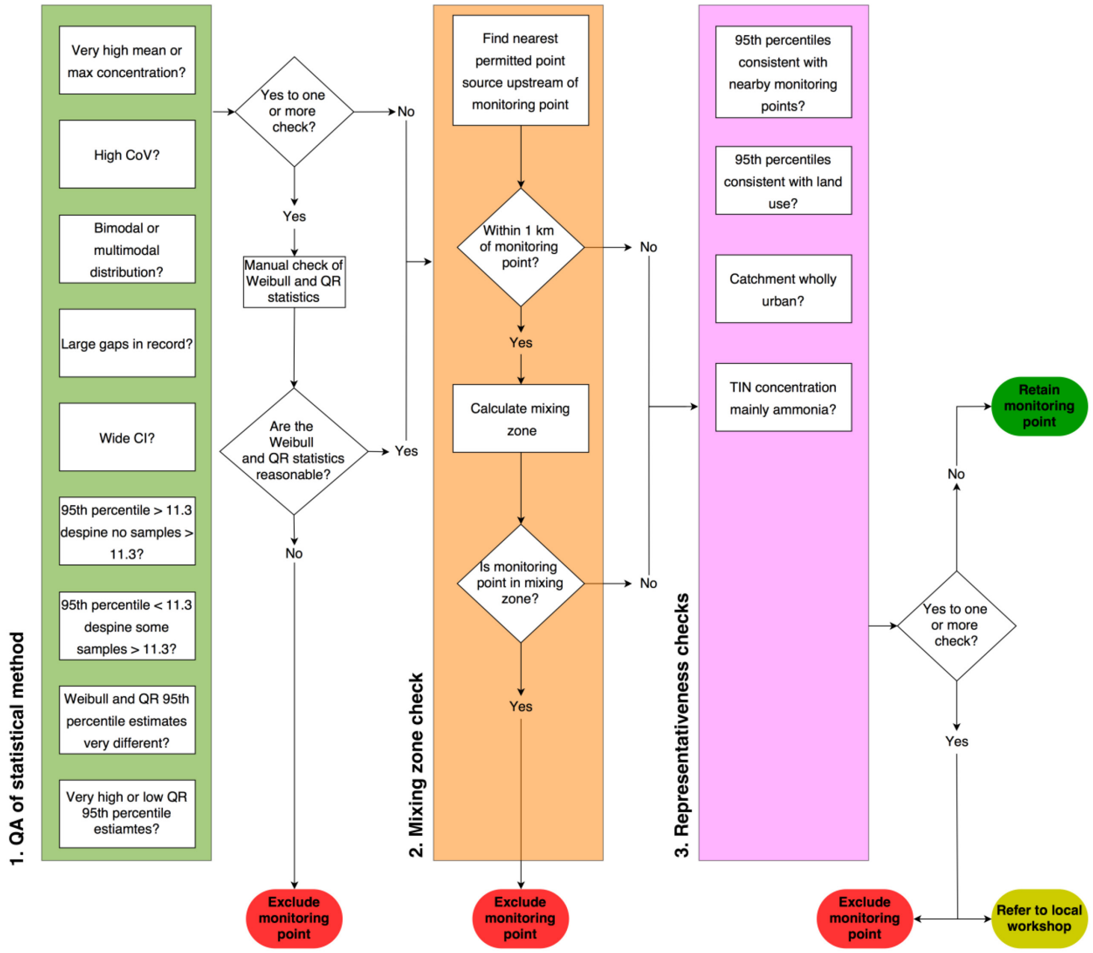


Figure 4.4 - Flow chart describing monitoring point level checks on the statistical methodology and the representativeness of the monitoring point. The decision to refer a monitoring point for discussion at a local workshop may be made at any point.

Case Study

Monitoring point 22244040 (River Salwarpe at Chapel Bridge, Droitwich) on the River Salwarpe, a tributary of the River Severn, has been in an NVZ designation since 2002. The River Salwarpe joins the Severn just upstream of Worcester. As seen in Figure 4.5, it has had consistently high TIN concentrations over the last 25 years.

From 1990 to 2014 this monitoring point has 216 TIN samples.

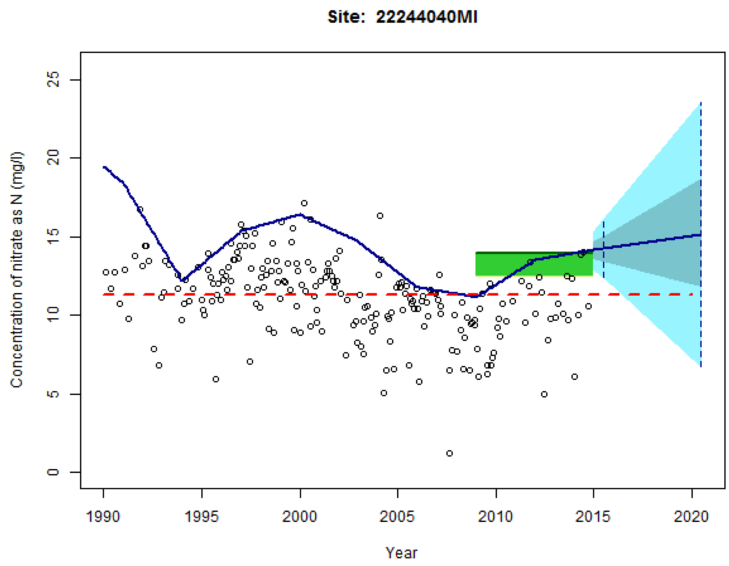


Figure 4.5 - TIN as N concentrations over time at sample point on the River Salwarpe

The monitoring point was sampled every year between 1990 and 2014. With 216 samples and 6 years of data between 2009 and 2014, the Weibull method was used to calculate current TIN and the quantile regression method to calculate future TIN.

The current TIN as N 95th percentile is 13.85 mgN/l and the lower 90th confidence interval is 12.43 mgN/l. As the lower 90th confidence interval exceeds 11.3 mgN/l, the sample point is assigned a class of 6 for current TIN.

The future TIN as N 95th percentile is 15.13 mgN/l, the lower 90th confidence interval is 6.72 mgN/l and the lower 50th confidence interval is 11.81 mgN/l. As the lower 90th confidence interval does not exceed 11.3 mgN/l but the lower 50th confidence interval does, the sample point is assigned a class of 5 for future TIN.

As the current TIN class is greater than the future TIN class this monitoring point is assigned an overall monitoring class of 6.

Figure 4.5 shows the final TIN trend graph and here are the results of the various checks we carry out for this monitoring point before concluding we can use it in our assessment.

1. The maximum and mean average TIN are 17.2 and 11.1 mgN/l respectively, so are not unusually high.
2. The current and future TIN predictions are 13.9 and 15.1mgN/l respectively and therefore not usually high.
3. The coefficient of variation is 0.22. This is quite normal for environmental data.
4. There are no obvious signs of bimodal or multi-model distribution.

5. There are no gaps in the data, with samples taken in all years between 1990 and 2014.

6. 110 samples exceed 11.3 mgN/l between 1990 and 2014, including multiple exceedences in the current review period.

7. The current 95th percentile TIN does not vary by much depending on the analysis method (13.9 mgN/l using the Weibull method, 14.2 mgN/l using the Quantile Regression method).

8. The nearest discharge is a private discharge for a hotel, which is 1.1 km upstream. The nearest upstream water company discharge is over 4km upstream. The river width at the monitoring point is 8m wide so the monitoring point is beyond the mixing zone for either discharge.

9. The current 95th percentile TIN at the upstream monitoring point is 19.4 mg/l and the current 95th percentile TIN exceeds 11.3 mgN/l at all other monitoring points in the WFD river catchment. The current TIN estimate here is consistent with nearby monitoring points.

10. The catchment is almost exclusively arable, improved pasture and urban, so the high concentrations are consistent with the highest risk land uses in the catchment.

11. The mean average percentage contribution of Ammonia to TIN samples taken between 1990 and 2014 is 1.2%, i.e. not significant.

5. Land Use Modelling

We use statistical modelling, alongside the analysis of monitored water quality data to assess whether a WFD river catchment is polluted or at risk of becoming polluted. We call these statistical models the NVZ land use modelling and they are key to the surface water method because they:

- provide an assessment of the risk of pollution in water bodies that don’t contain monitoring sites

- they tell us what the TIN should be based on the land use in a particular catchment and therefore provide context to the monitoring data.

Two models have been developed to predict TIN concentrations in main rivers and in tributaries/headwaters, referred to as the catchment and water body models respectively.

- The catchment models use data from the entire upstream area and predict the concentration at the bottom of the main river.

- The water body model uses data from individual water bodies only and the prediction is for the local WFD river catchment only.

The catchment models provide evidence towards designation of catchment NVZs, whereas the water body models provide evidence towards designation of water body NVZs.

5.1 Data

Response variables

The response variables are what we are attempting to predict and in this case are the current Weibull and quantile regression 95th percentile estimates of TIN concentrations at individual river monitoring points, as described in section 4.

For the catchment model, 95th percentile TIN from main river sample points are used as the response variable. For the water body model 95th percentile TIN from tributary and headwater monitoring points are used as the response variable. Where there is more than one sample point per water body classed as main river or tributary and headwater, a mean 95th percentile is used.

Explanatory variables

Datasets are collated to derive potential variables for explaining variation in observed TIN concentrations (the response variable) and subsequently predict TIN concentrations for unmonitored water bodies. They represent many of the known sources and controls of nitrogen pollution to the environment:

- diffuse agricultural sources
- urban sources
- point sources (agricultural, industrial, landfills, quarries, waste water and other sources)
- septic tanks
- historic landfills
- atmospheric deposition.

Diffuse agricultural sources

We use the NEAP-N model to represent the losses from agricultural land. The NEAP-N model (Anthony et al., 1996; Lord and Anthony, 2000; Silgram et al., 2001) was developed under Defra Water Quality funding as a policy tool to allow estimation of TIN losses from agricultural land, applicable to any catchment in England and Wales. It is acknowledged as the foremost national nitrate leaching model in the UK and is widely used throughout the UK and in pan-European studies. It has been an important part of the last three NVZ reviews.

NEAP-N builds upon previous models and field experiments and is essentially an export coefficient model that considers land use, climate and soil type. Nitrate loss potential coefficients are assigned to each crop and livestock type and the model summarises these to predict a total annual loss from agricultural land, which can be split into losses of nitrogen and nitrates from:

- fertiliser applications and losses due to other cropping practices on arable land
- manures from housed animals applied to arable and managed grassland
- excreta from grazing livestock to managed grassland and rough grazing land
- atmospheric deposition on arable land, managed grassland, rough grazing land, woodland and open water

NEAP-N includes a water balance model and a leaching algorithm, which calculates the proportion of the potential loss that is actually leached.

The input data to NEAP-N are the agricultural census (for each of the cropping and livestock categories); the dominant soil type (from the Cranfield University NatMap soils dataset); mean annual rainfall (from the UKCP09 1961-90 baseline climate dataset); and potential evapotranspiration for the different crop types.

The version of NEAP-N used in this project has been modified to better represent the impacts of atmospheric deposition (Lord et al., 2007) and the land use data in the model is based upon results from the 2014 June agricultural census (the most contemporary data available). For more details on this version of the NEAP-N model, please refer to the latest report (Lee et al 2015).

We do not use NEAP-N to **directly** predict nitrate levels in surface waters because:

- The output of the model is total annual nitrate-N loss from the soil profile for agricultural land. It makes no assumptions about the hydrological pathways that the nitrates then follow, either to groundwater or via surface flow pathways.
- It also does not consider the effects of de-nitrification once the nitrates are mobilised from the soil or the effects of previous losses in the groundwater returning to the surface.
- It does not consider any nitrate losses from non-agricultural land.

- The land use data is contemporary but the rainfall data that leads to the loss of nitrogen from the soil is based on the long-term average. It will therefore not consider the effect on nitrate-N losses if the year in question was particularly wet or dry.

- The model predicts the loss from the crops and livestock in the area but does consider how these change due to current mitigation options (such as NVZ Action Programme measures or Catchment Sensitive Farming).

For these reasons, we use NEAP-N as part of the evidence supporting NVZ designations and interpret it alongside other data.

The NEAP-N model predicts nitrogen losses on a $1km^2$ grid in kg ha^{-1} yr^{-1}. This is converted to kg km^{-2} yr^{-1} and losses are aggregated to WFD river catchments assuming that nitrate losses are uniformly spread across each grid square and then apportioning the losses to each WFD river catchment based on proportional overlap. Losses are then aggregated upstream so that total and broad land cover category variables are generated for the catchment and waterbody models.

The losses as loads are converted to concentrations to standardise across differences in water body or catchment size and rainfall regime. Two methods of standardisation are used; flow and hydrologically effective rainfall (HER). Flows are taken from the Low Flows 2000 models and the CERF model and HER is derived by the CERF model (Griffiths et al. 2008; Holmes et al. 2005).

Urban sources

Diffuse urban nitrogen losses were estimated using the method provided by Mitchell (Mitchell 2001, 2005, 2006). The Mitchell method generates a loss from the following urban land uses:

- urban green space
- industrial
- commercial
- residential (including loss from leaking water mains and sewers)
- roads
- construction and urban population

Input data are derived from the 2006 CORINE (EEU 2007) land cover dataset and aggregated to WFD river catchments.

Export coefficients are expressed as an event mean concentration per urban land use. In order to integrate the urban losses with other sector losses we convert this to a load in kg/yr. This is achieved using the CERF long term average (1971 - 2000) HER. The HER arising from each urban land cover is multiplied by the event mean concentration for that land cover class. The loads for all land cover classes are combined into a single urban diffuse loss per 1km grid square. A second version assuming that only HER which flows via surface pathways, as indicated by the HOST Base Flow Index (BFI), drives urban losses was also produced (Boorman, Hollis, and Lilly 1995).

The 1km^2 losses are aggregated to water bodies and accumulated upstream and then standardised by flow and HER as described in the agricultural sources section.

Point sources

Several point source explanatory variables are derived from the Environment Agency WIMS database. Information on discharge location, consented flow and consented ammoniacal nitrogen, nitrate and nitrite concentrations are extracted for all discharge consents with numerical conditions.

Where a STW does not have a consent for nitrogenous compounds, a measured value for that STW is used if available, otherwise the average concentration for all STWs with measured data is used.

Sites with a dry weather flow consent condition had the dry weather flow value multiplied by 1.3 so that they are comparable with consents with a mean flow consent condition and all flow consent values are converted to Ml d^{-1} (Pygott 1999).

Nitrate, nitrite and ammoniacal nitrogen loads are generated by multiplying the consent flow condition by the quality consent condition. The total point source flow and load are calculated per waterbody using the discharge location information from WIMS. These are accumulated upstream using the same method as above. The load estimates are converted to concentration estimates using accumulated flow consent conditions.

Septic Tanks

The term septic tanks is used here to refer to any private, residential sewage disposal system, the majority of which will be septic tanks. An estimate of the total number of septic tanks per waterbody was made based on the likelihood of individual properties being connected to the sewerage network. A nitrogen loss to the environment connected to a septic tank of 1.64 kg year^{-1} person^{-1} is multiplied by an estimate of the population connected to septic tanks to give the total load of nitrogen released by septic tanks per water body per year (Carey and Davidson 2010, Burgess 2013).

The estimated septic tank load is split into the load assumed to flow to surface water bodies and the load to groundwater bodies using the HOST BFI (Boorman, Hollis, and Lilly

1995). Loads are accumulated upstream and then standardised by flow and HER as described in the agricultural sources section.

Historic landfills

Historic landfills with buried biodegradable material may be significant local sources of ammonia. Explanatory variables are generated using Environment Agency spatial data on the location, extent and composition of historic landfill sites. The total area per WFD river catchment, percent coverage per water body and a binary factor indicating if a WFD river catchment contains a historic landfill with biodegradable material are derived.

The total area and biodegradable material indicator are accumulated upstream as per the agricultural sources sector and the percent coverage is re-calculated for the entire upstream contributing area.

Base Flow Index

BFI is not a source of nitrate, but is included in the model to test whether flow variability (a low BFI catchment should be expected to have a more variable flow regime) affects 95th percentile TIN concentrations.

This analysis uses the BFI from the NSRI NATMAP 1000 dataset, that is part of the Hydrology of Soil Types (HOST) dataset (Boorman, Hollis, and Lilly 1995).

The HOST 1km^2 BFI is aggregated to WFD river catchments but converted to an area weighted average rather than accumulation by summing.

5.2 Modelling method

For the catchment model, waterbodies and their upstream areas smaller than 20km^2 are excluded and for the waterbody model waterbodies smaller than 20km^2 are excluded. The 20km^2 threshold is used as the aggregation process for generating 1km^2 landuse datasets is unreliable for smaller areas. For the catchment models, 2,176 waterbodies are included (34% of waterbodies, covering 64% of the land area) and for the waterbody model 509 waterbodies are included (8% of waterbodies, covering 22% of the land area).

The predictions made using the final models are converted to broad risk scores in the final methodology so the designation of NVZs is fairly insensitive to the exact modelling method (EA 2016).

The distributions of the agricultural and urban losses across water body areas, and of the 95th percentile TIN concentrations across monitoring sites, are all highly skewed. To ensure normal residuals, data for model variables are transformed using logs to base 10 (log10(0.1 + x)) for both the response and selected predictor variables.

The model formulation is an ordinary least squares regression, assuming a normal error distribution. A single model is generated to cover all of England. The R statistical programming environment is used to perform the modelling (R Core Team 2015).

Building the models

The explanatory variables listed in Table 5.1 are used in the model selection process. A number of combinations are manually tried with no hard rules on which variables will be considered in the same model; instead a common sense approach to compatible variables is used. By common sense we mean variable combinations are combined to avoid co-linearity and to avoid choosing more than one variable representing each source of TIN - for example not using both the Lerner and Mitchell urban diffuse N sources in the same model.

Model selection is performed by 5 fold cross validation with 10 repetitions using the R function cvTools::cvFit() (Alfons 2012). The preferred model is selected using root mean squared prediction error (RMSPE) as the cross validation cost term.

The preferred models use total agricultural NEAP-N load as opposed to the separate arable, grassland and rough grazing components. While the model tested using the individual components returned a lower RMSPE the coefficient for rough grazing was negative suggesting that rough grazing served to remove nitrogen from the system. This is demonstrably erroneous and a likely artefact of a strong inverse correlation between rough grazing and arable land. For the agricultural source data the HER standardised concentration is the preferred explanatory variable.

The urban load predictions as standardised by HER performed best in terms of cross validated RMSPE and model fit. The urban load by all flow pathways out-performed the urban load via surface pathway for both the waterbody and catchment model.

Total point source flow out-performed point source load and concentration in terms of fit and cross validated RMSPE for all models.

The historic landfill variables did not improve the models in terms of model fit or cross validated model performance and are therefore omitted from the preferred models. Similarly BFI did not improve the models in terms of fit or cross validated RMSPE and was omitted.

Table 5.1 - Parameters tested in model development

Response variable	Log10(95^{th} percentile TIN)	
Explanatory variables	BFI	Base Flow Index (unitless)
	Log10(0.1 + Total NEAP-N load) Log10(0.1 + Arable NEAP-N load Log10(0.1 + Grassland NEAP-N load) Log10(0.1 + Rough grazing NEAP-N load)	Arable sources (all in kg yr^{-1} & mg l^{-1})
	Log10(0.1 + Lerner Urban load A (no population estimate)) Log10(0.1 + Lerner Urban load B (includes population estimate)) Log10(0.1 + Mitchell Urban load C (Via surface water pathways only)) Log10(0.1 + Mitchell Urban load D (Via all pathways))	Urban sources (all in kg yr^{-1} & mg l^{-1})
	Log10(0.1 + Point source flow) Log10(0.1 + point source ammonia) Log10(0.1 + point source TON) Log10(0.1 + point source count)	Point source (flow in Ml yr-1, ammonia and TON in kg yr^{-1} & mg $l^{-1)}$
	Log10(0.1 + septic tank load to river)	Septic tanks (in kg yr^{-1} & mg $l^{-1)}$
	Log10(0.1 + Landfill area) Log10(0.1 + Landfill coverage)	Historic landfills (in km^2 and % of total land area respectively)

Table 5.2 shows the final preferred models.

Table 5.2. - Preferred models

Model	Response	Explanatory variables	Cross validated RMSPE	% variance accounted for
Catchment	Main river Log10(95th percentile TIN)	Log10(0.1 + Total NEAP-N load) + Log10(0.1 + Point source flow) + Log10(0.1 + Mitchell Urban load D + Log10(0.1 + septic tank load to river)	0.243	70%
Waterbody	Tributary and headwater Log10(95th percentile TIN)	Log10(0.1 + Total NEAP-N load) + Log10(0.1 + Point source flow) + Log10(0.1 + Mitchell Urban load D + Log10(0.1 + septic tank load to river)	0.304	57%

5.3 Model checking

For both the waterbody and the catchment models, the residuals are normally distributed and have constant variance. There is some structure in the residuals (over predictions at high and low values and under predictions for mid-values) indicating some non-independence. When the residuals for both models are spatially plotted, the catchment model slightly over predicts N concentration in Norfolk and Suffolk. For the waterbody model there is no spatial structure in the residuals.

The over-prediction in Norfolk and Suffolk may be due to an over estimation in the NEAP-N total load coefficient (NEAP-N total load is highest in East Anglia). However the lack of corresponding over prediction in other areas of high NEAP-N total load (Lincolnshire, East Riding) suggests the effect of any bias is small. No corrections to the models were made after inspection of the residuals.

5.4 Model use

The models are used to make estimates of 95th percentile TIN in all waterbodies larger than 20 km^2. Estimates of upper and lower 95th and 75th prediction intervals are also made to allow each waterbody to be given a catchment and waterbody modelling risk score that becomes a line of evidence for that WFD river catchment in the evidence matrix (described in section 6).

Figures 5.1 and 5.2 show the catchment and waterbody model fits. Figures 5.3 and 5.4 show the catchment and waterbody model 95th percentile TIN predictions.

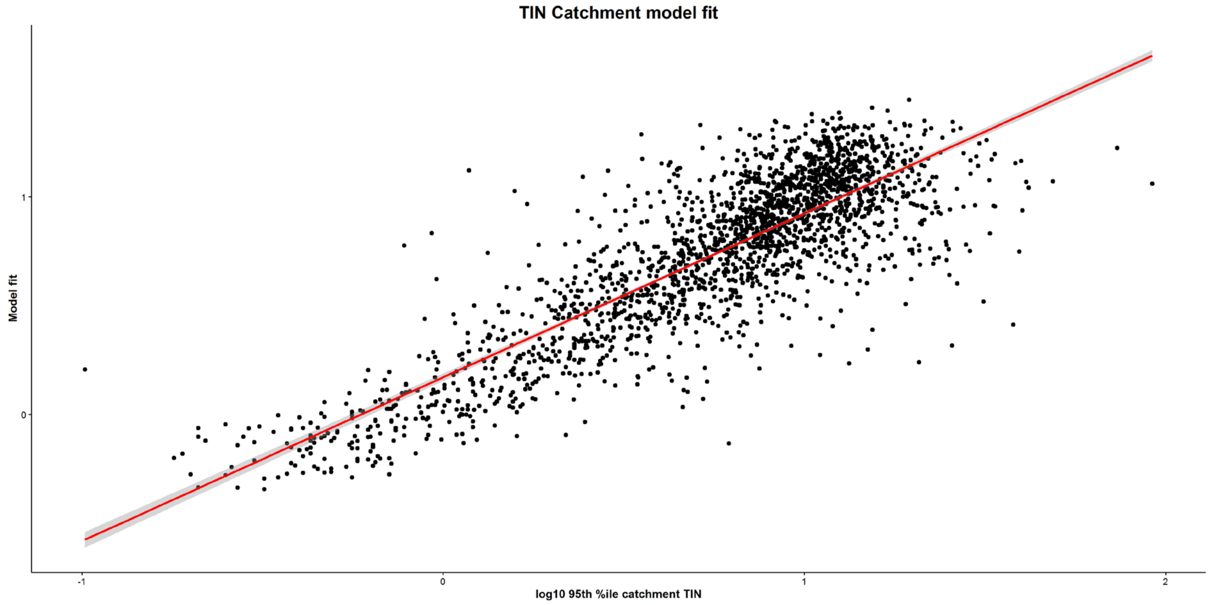


Figure 5.1 - Catchment model fit, in log space

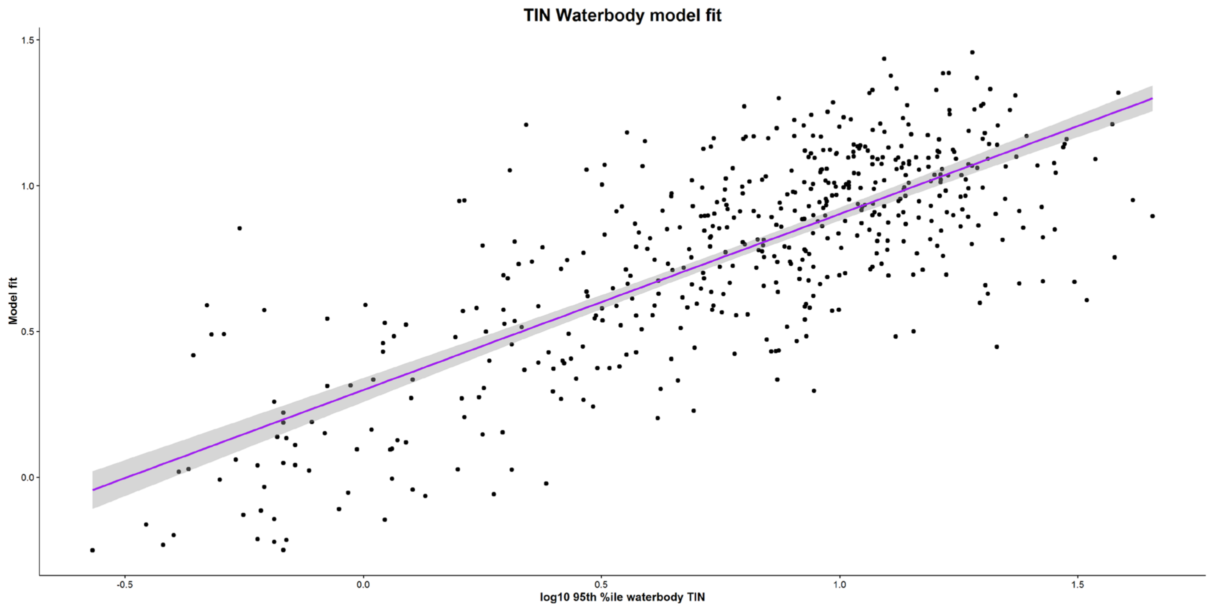


Figure 5.2 - Waterbody model fit, in log space

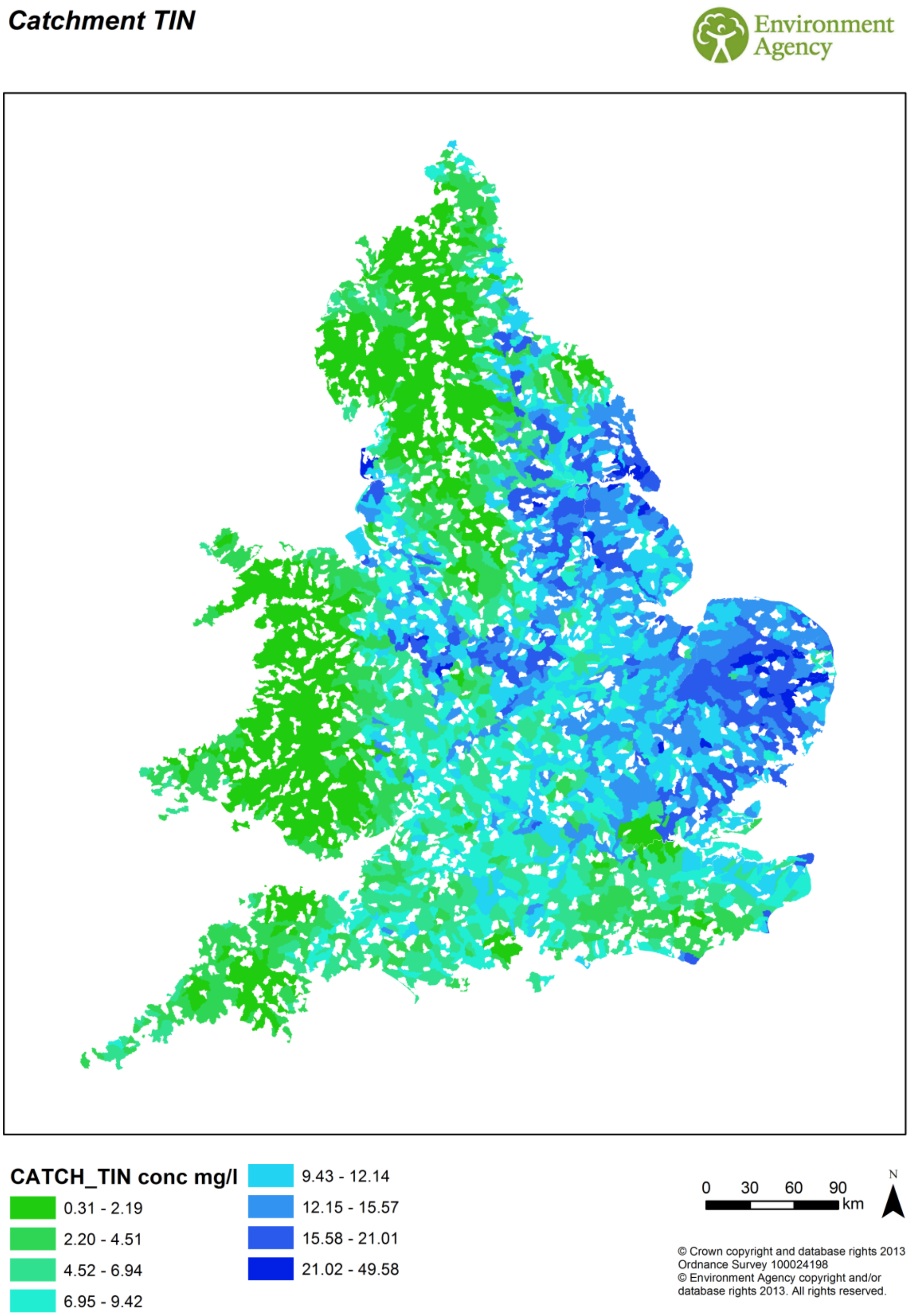


Figure 5.3 - Catchment model 95th percentile TIN predictions

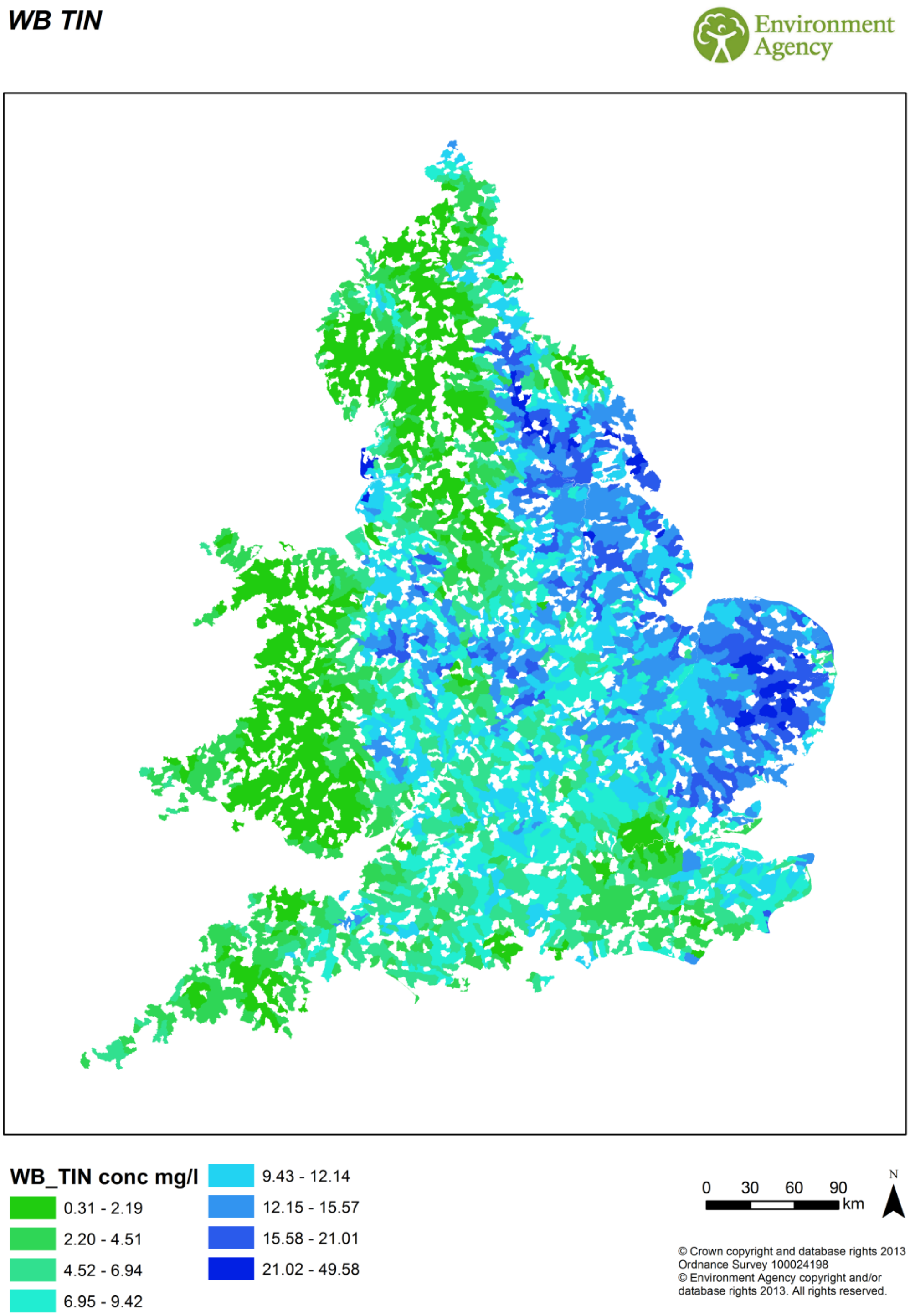


Figure 5.4 - Waterbody model 95th percentile TIN predictions

6. Combining Monitoring and Modelling results in the surface water evidence matrix

This section describes the process of bringing together the monitoring and modelling evidence described in sections 4 and 5, and using them to update our assessment of the weight of evidence supporting NVZ designations in the current review.

6.1 Lines of evidence

Each WFD river catchment is assessed twice; once to assess the evidence supporting a catchment NVZ designation and once to assess the evidence supporting a water body NVZ designation. Not all WFD river catchments have sufficient monitoring data to have a monitoring class and some are too small to apply the land use model with confidence. In these circumstances, the monitoring or model class will be 0 (zero) indicating insufficient data.

Catchment NVZ designation lines of evidence

To assess the evidence supporting a catchment NVZ designation we combine:

- the worst main river monitoring class per WFD river catchment - this can be either a current or future monitoring class.
- the catchment land use model class for each WFD river catchment

Water body NVZ designation lines of evidence

To assess the evidence supporting a water body NVZ designation we combine:

- the worst tributary monitoring class per WFD river catchment - this can be either a current or future monitoring class
- the water body land use model class for each WFD river catchment

All four lines of evidence (catchment monitoring, catchment modelling, water body monitoring and water body modelling) are scored on the 1 to 6 scale, with 1 being a confident pass and 6 being a confident fail:

Table 6.1 - Monitoring and model classes

Monitoring / modelling class	Definition
1	Confident pass
2	Moderate pass
3	Face value pass
4	Face value fail
5	Moderate fail
6	Confident fail
0	Insufficient data

6.2 The evidence matrix

The evidence from the monitoring data and the land use model is combined to assess the strength of evidence for TIN pollution in each WFD river catchment. A WFD river catchment is considered to be polluted if either the water body or the catchment tests provide evidence of pollution. The evidence matrix is exactly the same as the one used in the 2013 review.

Figure 6.1 shows the evidence matrix used to determine water bodies with TIN pollution. Each combination can be grouped into one of four outcomes:

- Red cells indicate that the water body has sufficient evidence to be assessed as polluted or at significant risk of pollution.
- Green cells indicate that the water body is assessed as not polluted but cannot be classed as at low risk of pollution either.
- Blue cells indicate that the water body is not polluted and that the risk of TIN pollution from agriculture is low. This category is used to help determine whether or not existing NVZs should be removed from designation (see section 7).
- Grey cells indicate water bodies that cannot be assessed because of a lack of evidence: they contain no monitoring sites and lack a reliable prediction from the land use model.

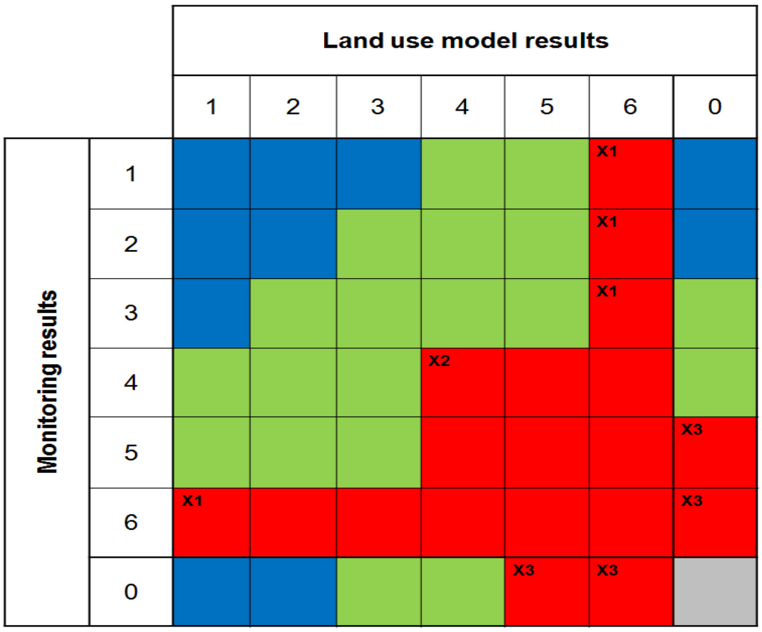


Figure 6.1 - Evidence matrix for surface water NVZ designations

The evidence matrix is used as a guide to decide whether or not surface waters are polluted. Discretion is exercised in all cases, but particularly in cells marked “X...”, where:

X1: the monitoring data and land use model give contradictory results

X2: the evidence for pollution carries a low degree of confidence

X3: the decision is informed by only one strand of evidence (either monitoring data or the land use model prediction)

In all cases, there is a presumption in favour of designation but local knowledge and detailed data quality checks provide an additional level of scrutiny and inform the final decision. Section 8 gives further details of the type of factors taken into consideration at the local workshops.

6.3 Converting evidence matrix scores into NVZ areas

Having identified WFD river catchments requiring catchment NVZ designations and those requiring water body NVZ designations, we apply the simple rules described in section 2.5 earlier. These rules state that:

- Water body NVZ designations result in the individual WFD river catchment becoming an NVZ
- Catchment NVZ designations result in all WFD river catchments upstream and including the ‘designating’ WFD river catchment being merged into a single NVZ

The only exceptions to this rule are catchment NVZs based on the land use model only. In such cases the normal catchment NVZ rule applies until we reach an upstream WFD river catchment with a catchment monitoring score. If that score does not indicate a catchment failure then the NVZ stops at this WFD river catchment. If it does indicate a catchment failure upstream, then a new NVZ starts from this WFD river catchment. We start a new NVZ here to show that this area meets different designation criteria.

6.4 Case study: applying the evidence matrix to the Essex Stour catchment

Once more it is useful to see the application of the method using actual data and real places. Figure 6.2 below shows the Stour catchment on the Suffolk/Essex border.

The dots represent the monitoring data and the background areas are shaded to represent the land use model predictions for the individual WFD river catchments.

Each has a score of 1-6 that translates to the confident pass - confident fail scale the we describe in table 6.1 above) We combine the worst monitoring point with the land use model result and see which combinations trigger designation rules.

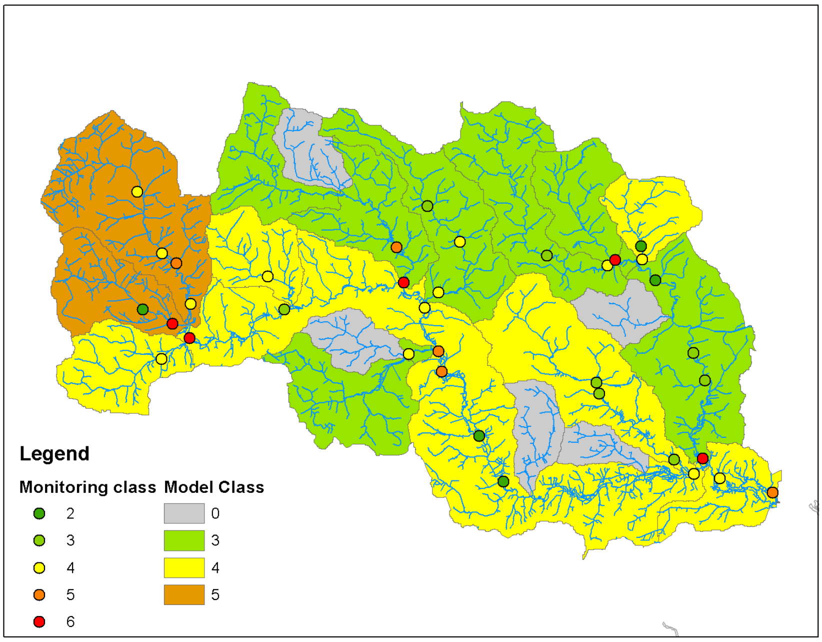


Figure 6.2 - Monitoring and modelling classes for the Stour catchment

In the Stour catchment we see the worst monitoring class varies between moderate pass (2) and confident fail (6), with failing monitoring points throughout the catchment. The most common land use model classes are face value passes (3) and face value fails (4) but the model predicts moderate fails (5) in the headwaters. Where the monitoring is a class 6, a confident fail, these would also qualify for designation. All other areas need to be assessed according to the combination of monitoring and model classes.

In Figure 6.3 below, we see which WFD river catchments we designate using the evidence matrix to each unit in isolation. The hatched areas are those water bodies where designation is triggered. Individually, there are a number of WFD river catchments that we designate, three that require further investigation (hatched in Figure 6.3) and 10 WFD river catchments that are designated based on the evidence for individual WFD river catchments. The most important WFD river catchment in the Stour, however, is the one at the catchment outlet.

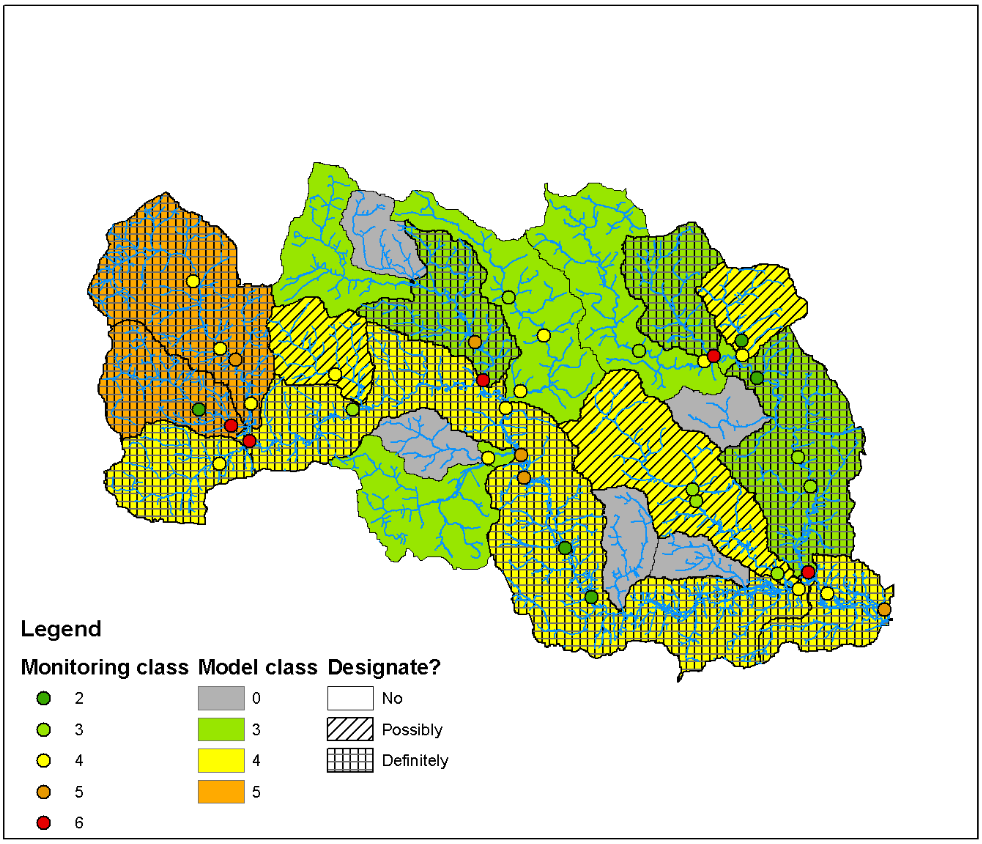


Figure 6.3 - Applying the evidence matrix to individual WFD river catchments in the Stour catchment

Figure 6.4 shows what happens when we differentiate between catchment and waterbody NVZs. The combination of main river monitoring and catchment model failures require us to designate all land draining to that unit. The WFD river catchment at the outlet is a moderate fail (class 5) for the main river monitoring and a face value fail (class 4) for the catchment model so all land upstream is designated, as shown by the cross-hatched area.

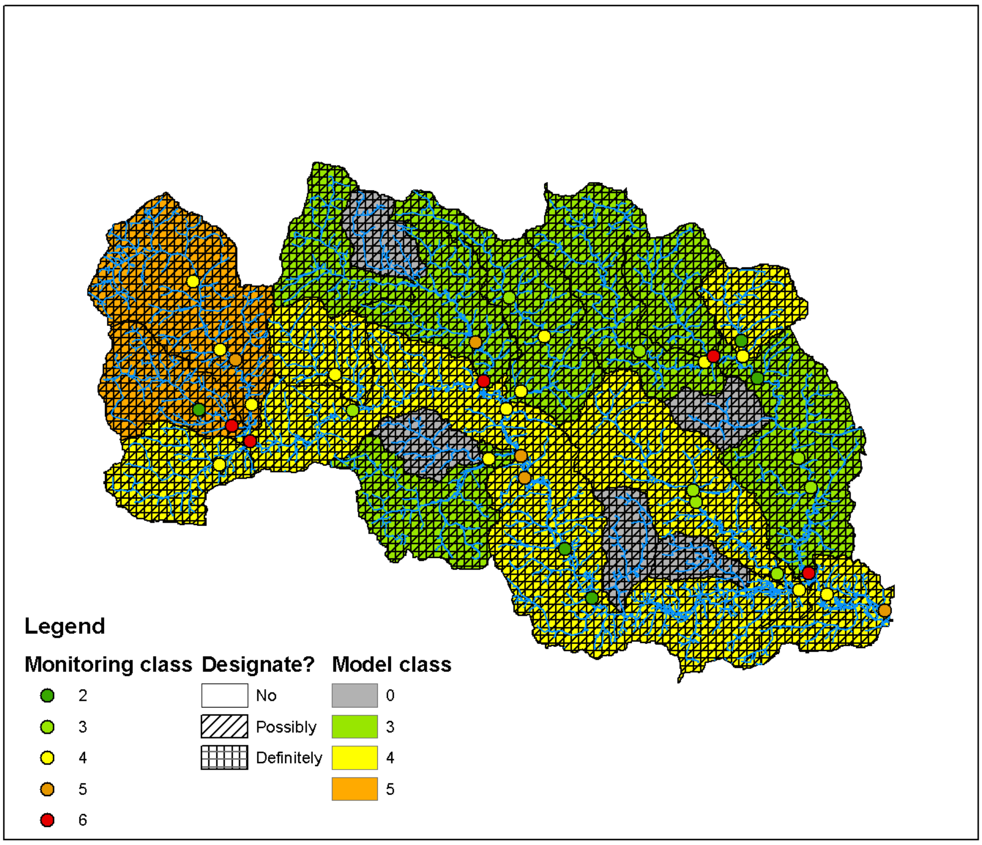


Figure 6.4 - Example of applying the evidence matrix and defining the NVZ based on the catchment NVZ designation in the Stour catchment

7. De-designation

The potential removal of an existing NVZ designation needs to be considered from a clear understanding of the Directive requirements and the risks of removing the control measures. Furthermore, if land is removed and water quality subsequently deteriorates then we will be required to designate the land again. This could undermine the NVZ process and have significant cost implications for farmers who remove any infrastructure associated with the NVZ Action Programme only to be required to reinstate it.

However, under certain circumstances, notably existing NVZs where water quality has improved sufficiently and agricultural land use, in combination with other TIN sources, no longer poses a significant risk of pollution, it is appropriate to remove NVZ designations.

In recent designation rounds, no NVZs have been removed as the designations had not been in place long enough to meet the criterion that water quality improvements have to be sustained for at least two successive designation rounds.

In 2013, the MRG agreed the broad criteria for identifying areas where de-designation should be considered:

1. monitoring data demonstrates that the water bodies are no longer polluted

2. agricultural land use is low risk as a source of TIN

3A. the improvement in water quality is sustained over at least two NVZ reviews

3B. the likely cause(s) of the water quality improvement can be determined

4. Land previously identified as draining to a polluted water can, on the basis of improved data, reasonably be determined to not drain to a polluted water

The 2017 review is the first time that the method can actually lead to de-designation. In applying the broad criteria from 2013, we made refinements and developed detail for aspects that were unclear.

This provides greater clarity over the circumstances in which removing a designation may be appropriate; addresses the outlined risks; and adds a further criterion to cover the risk to WFD objectives. To remove a previously designated area, **all four criteria must be met, with the weight-of-evidence considered on a case by case basis.**

1 Monitoring data demonstrate that the waters are no longer polluted

To meet this criterion, there must be a low risk of exceeding the NVZ threshold (i.e. one of the blue squares in the matrix shown earlier in Figure 6.1) for two consecutive reviews. It is important to recognise that a monitoring or modelling score of 3 means that the 95th percentile TIN concentration could be 11.29 mgN/l and we have only slightly more confidence that it will not exceed the Nitrates Directive standard than that it will. In other words we do not believe that a score of 3 should, on its own, be a potential trigger for de-designation because we cannot be confident that a small change in the catchment will not result in a reversion to polluted water status.

The Directive provides an indication in Article 6 as to ‘low risk’ waters, based on no samples exceeding 25 mg/l as nitrate (or 5.65mgN/l) over the last 8 years. This lower value could be used in order to reduce the possibility of zones changing between designation, de-designation and back to designation in consecutive cycles. However, that would be a significant change to the published method and thus beyond the scope of the 2017 review. We therefore retain the published criterion but add the following:

If the evidence matrix indicates there is low risk of the water being polluted for two consecutive cycles, we consolidate this by checking to see if there are any TIN samples above 11.3 mgN/l during the last two cycles. If there are one or more TIN samples above 11.3mgN/l then we consider that the risk of pollution is still present and this is taken into account when considering the remaining criteria

2. Agricultural land use is low risk as a source of TIN

This criterion is designed to increase confidence that removing the NVZ Action Programme will not lead to water quality deteriorating. Typically it will be necessary to show that land use in the NVZ has changed to become a low risk source of TIN. If the Action Programme has played a significant role in the water quality improvement then it will be necessary to consider, case by case, whether removing it would result in TIN rising again. If it would then it is likely to be inappropriate to remove the NVZ designation.

We look at loading estimates (from the NEAP-N model) from 2000 and compare these with current estimates. This will reflect any changes to land use. If we see a significant drop in loading estimates over time then this is indicative of a decreased risk from agriculture. This alone, however, is not an appropriate test of the inherent risk from agriculture, as NEAP-N predictions are a snapshot of potentially available nitrogen based on crop and livestock data for a given year and the model takes no account of the effect of mitigation and land management. It is important to check, as far as is possible, that changes to land use and land management are permanent. The local workshops are used to help determine the risk from agriculture.

We do not de-designate an area if the reason for improvement in a catchment is considered to be the impact of the NVZ Action Programme and/or other agri-environment initiatives such as Stewardship or Catchment Sensitive Farming.

3a. The water quality improvement is sustained over at least two designation cycles

This is part of showing that the water is no longer polluted and that deterioration is unlikely. As outlined above, Criteria 1 should be met for at least two successive NVZ reviews. To increase our confidence that the improvement is sustained, we also look to ensure there is no recent upward trend in TIN concentration. Monitoring data, in particular, can be influenced by short-term changes in water quality and while the analysis may still classify the WFD river catchment as low risk, there is a significant difference in the confidence that this will be sustained in the two examples shown below.

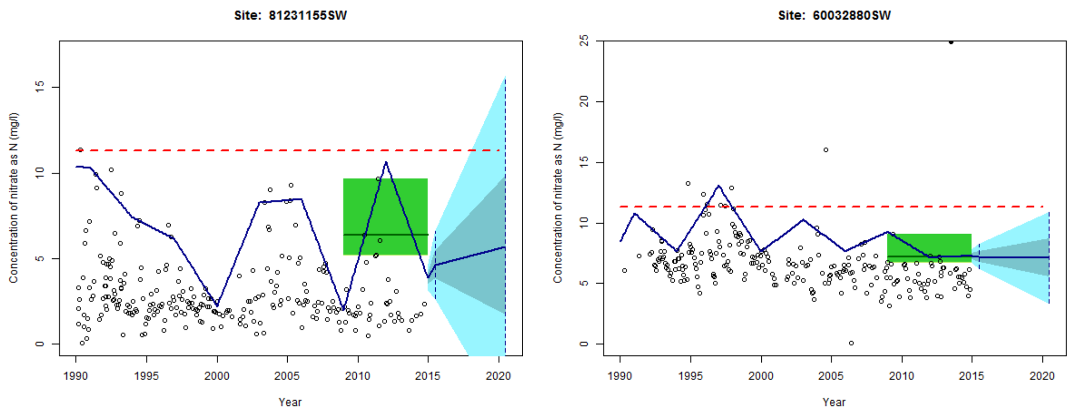


Figure 7.1 - Examples of varying recent trends in monitoring data

Both of the monitoring points shown in Figure 7.1 would score 2 and if combined by a score of 3 or less for the land use model would provide evidence that the water is no longer polluted. However, the variability and the recent upward trend in the graph on the left provide less confidence in the trend and this must be taken into consideration when we look at the evidence supporting de-designation as a whole.

3b The likely cause(s) for the water quality improvement can be determined

Establishing the cause of observed improvements in water quality is essential to demonstrate that nitrogen pollution will not return if an effective Action Programme is removed. Appropriate evidence is needed on a case by case basis. This is most relevant to circumstances where a dominant point source effluent discharge has been diverted or removed, leading to a significant improvement in TIN levels.

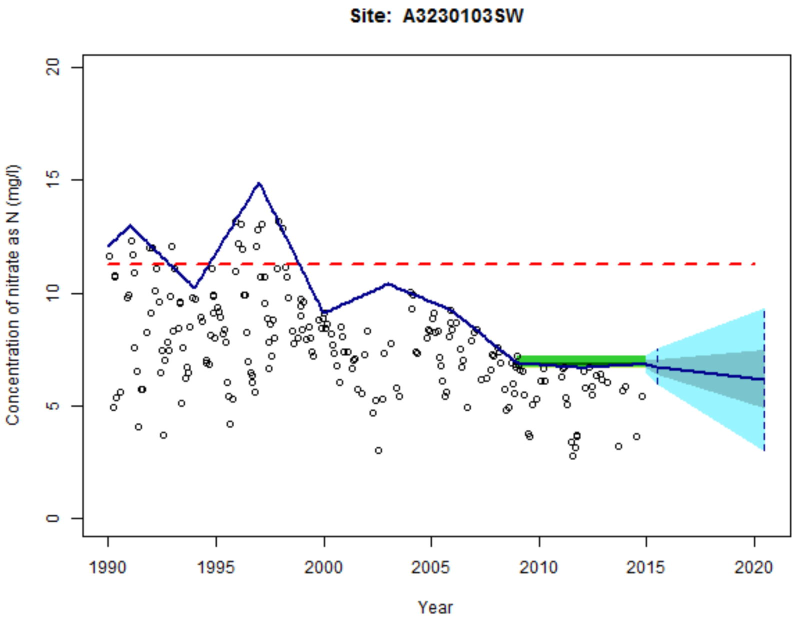


Figure 7.2 - Example of monitoring data in a catchment where the risk of pollution is likely to have reduced significantly

The graph above shows an example that illustrates the change we should see over time in order for de-designation to be confidently supported. The original designation was added in 2002 but since then there have been major improvements to the local sewage works resulting in TIN levels not exceeding 7 mgN/l as N since 2009 and not exceeding 11.3 mgN/l as N (50 mg/l as NO_3) since 1998. In this example;

- water quality has shown sustained improvement over two review cycles
- the 50mg/l nitrate threshold has not been exceeded in that time
- TIN levels are consistently well below the threshold so the risk of removing the Action Programme is low
- we can determine that the improvement is due to changes in the catchment that are independent of the Action Programme

All the criteria are met and if this is the sole failing monitoring point in the designating WFD river catchment then de-designation here is justified. We would ask the local Environment Agency teams whether **removal of the NVZ will compromise WFD objectives** because the implications of removing the NVZ designations should be considered in the context of the wider WFD Programme of Measures in each River Basin District.

In this review we have limited this to checking that de-designation would not have adverse affects on surface water Drinking Water Protected Areas where sites are at risk due to rising TIN or eutrophication problems where TIN is considered to be a causal factor. It could be argued that the benefits to WFD objectives of the NVZ Action Programme are relevant to more than just TIN and that we should check against other pressures as well. We consider this beyond the scope of the 2013 criteria and would require approval by a MRG.

8. Using local knowledge to test draft designations

The final step in the designation process is to consult local Environment Agency staff, as well as external stakeholders, and use their understanding of local pressures and water quality to test that the method can be appropriately applied in their areas. The forums for such discussions are a series of local workshops. By following the steps described in the previous eight sections of this document, we identify areas where the national lines of evidence indicate we should:

1. add a new NVZ area
2. extend/modify an existing NVZ
3. de-designate an existing NVZ, either partially or wholly

Any of these areas of potential change are the highest priority for discussion during local workshops.

In addition, existing NVZs are also prioritised for discussion if:

4. the monitoring and modelling lines of evidence are contradictory (as described in section 6)
5. national level investigations indicate that there could be issues with the monitoring points that caused designations in previous reviews (such as those described in section 4.2)
6. national evidence indicates that there has been a significant change in water quality based on the current review
7. our understanding of local hydrology has changed since the 2013 review

All the proposed 2017 surface water NVZs are identified as high, medium or low priority. High priority status is given to areas of change (numbers 1-3 above) and to queries of existing NVZs (numbers 4-7 above) about which we have concerns (as opposed to requiring clarification/confirmation). Medium priority status is given to existing NVZ areas where further clarification is preferable but we are confident that the NVZ is still justified. All other NVZs are considered low priority for discussion in local workshops.

Some of the most frequent discussion points are listed below:

1. Is there any reason to distrust the monitoring points we are using? Reasons include:

    a. Are we considering all the non-agricultural sources contributing to that catchment? Is there something that we have missed that unduly influences a monitoring point?

    b. Is there a tidal influence at that point? If there is tidal influence, however infrequently, then we are likely to exclude the point from our analysis.

    c. Is the monitoring point where we think it is? This is particularly important when determining whether it is a water body or catchment NVZ.

    d. Are we accurately representing the hydrological connectivity of the area? There are areas in England where WFD river catchments are the logical amalgamation of water bodies for reporting purposes but not necessarily for NVZ designations.

2. Can we explain any short-term ‘spikes’ or significant changes to the longer-term trends we observe in the monitoring data?

3. Is this an isolated NVZ and if so can we explain why TIN is higher here than the surrounding area?

4. Is there a history of pollution incidents that we should be aware of?

5. Are there planned changes, such as improvements to sewage works that will significantly improve TIN levels in the future?

6. Are there more detailed source apportionment studies that improve the evidence in this area?

The discussions at workshops are guided by questions about individual catchments raised during the designation process. The discussions are not limited to these questions but cover wider issues concerning the catchments. A summary of the workshop discussions is included in the data information for individual NVZs. If there are no such comments in the datasheet, this means that the NVZ was not discussed in detail.

It is important to stress that workshop discussions are part of the evidence for each NVZ and there are limits to what they can do. It can result in the following changes:

- If the local staff confirm that monitoring points are within the mixing zone of a point source, we remove this point from the assessment.

- If the workshop confirms that the flow from sewage works contributes over 80% of the flow at the failing monitoring point, we can remove this point from the assessment because it is not representative of water quality from the upstream area.

- If local staff confirm that the land use model does not accurately reflect land use, and the model is the reason for designation, then we consider over-riding the model in this area.

- If local staff support the view that a designation caused by a sharp future trend is unrepresentative of the water quality, then we consider over-riding the trend classification in these areas.

If the local recommendations result in the removal of a monitoring point or a score in the evidence matrix, this does not necessarily mean that we do not continue to propose the NVZ. In such circumstances, we re-run the evidence matrix scores for that area.

Local workshops cannot directly remove a NVZ because the local view is that agriculture is not the dominant source. We record recommendations from local teams that other sources are the reason for poor water quality, but unless those recommendations result in the exclusion of all failing monitoring and model data (based on the criteria outlined in sections 4 and 5 of this document) then we retain the designation.

The following section presents a case study of a fictional example to demonstrate the way a decision to designate or not designate would be made at a local workshop.

8.1 Case study: Discussing a new designation where the monitoring failure is downstream of a significant point source

One of the most frequent questions asked at local workshops is ‘How much is that STW contributing to the failure?’ As described in section 2.7.2, while we can calculate the contribution of a particular point source to the load of N at a monitoring site, the Directive gives no guidance on what the proportion of N from agriculture should be for a failing monitoring location to result in designation, and in fact does not refer to ‘source apportionment’ or the ‘significance’ of agricultural contributions at failing monitoring locations.

In this example, the resulting NVZ is a large main river catchment (120 km^2) which is a new failure, and which would designate an upstream catchment if accepted as an NVZ. There are two monitoring points within the catchment, both of which have deteriorated in quality since 2013, with the lower sample point deteriorating from a class 4 (face value fail) in 2013 to a class 6 (confident fail) in 2017. The predicted agricultural class has also declined from a 3 in 2013 to a 4 in 2017. So in 2013, both monitoring points were a matrix pass, but in 2017, the downstream monitoring location is a matrix fail. As this is a potential new designation, the catchment would be discussed at the local workshop.

The local staff explain that there have been two important changes in the catchment since 2009, the first that the STW had been expanded and rebuilt in 2011 and is now accepting waste from two smaller STWs that have been closed, and high nitrate waste from an animal feed factory that has surrendered its own consent. The second was that local farms have been converting from dairy to arable, partly because of economic conditions and partly because of the expansion of the animal feed factory. Figure 8.1 illustrates the failing monitoring point.

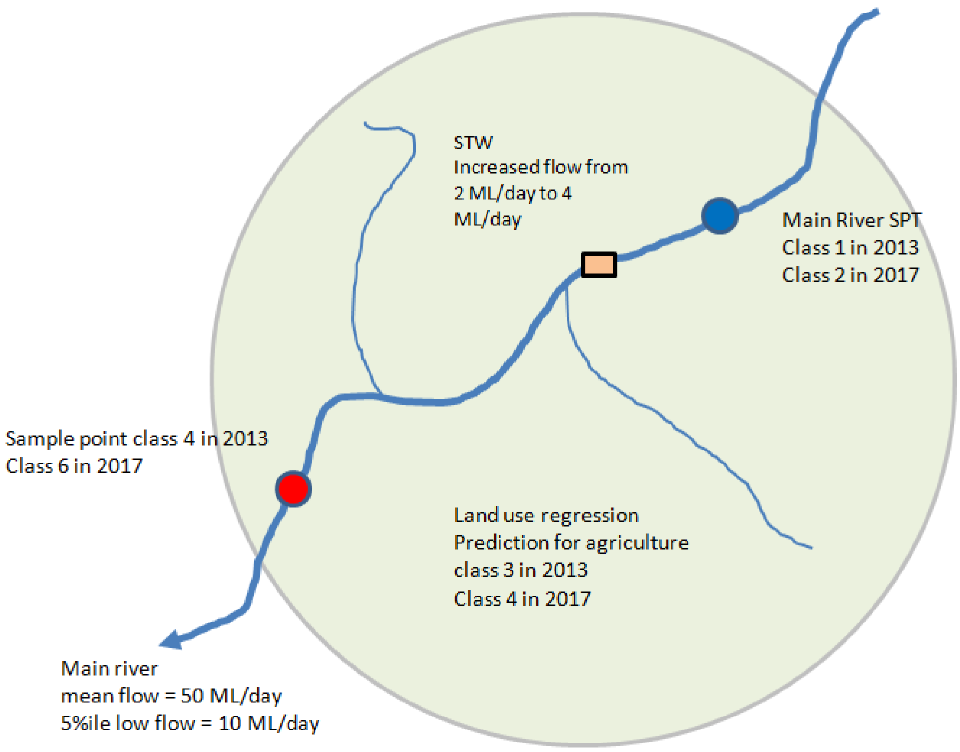


Figure 8.1 – Example of a failing monitoring point downstream of a sewage treatment works

The time series of TIN levels at the failing monitoring site shows that there is an upward trend at this site, and that the first failing sampling occasions were in 2009, before the expansion of the STW. The TIN levels at the site were already a face value fail in 2013. This is shown in Figure 8.2.

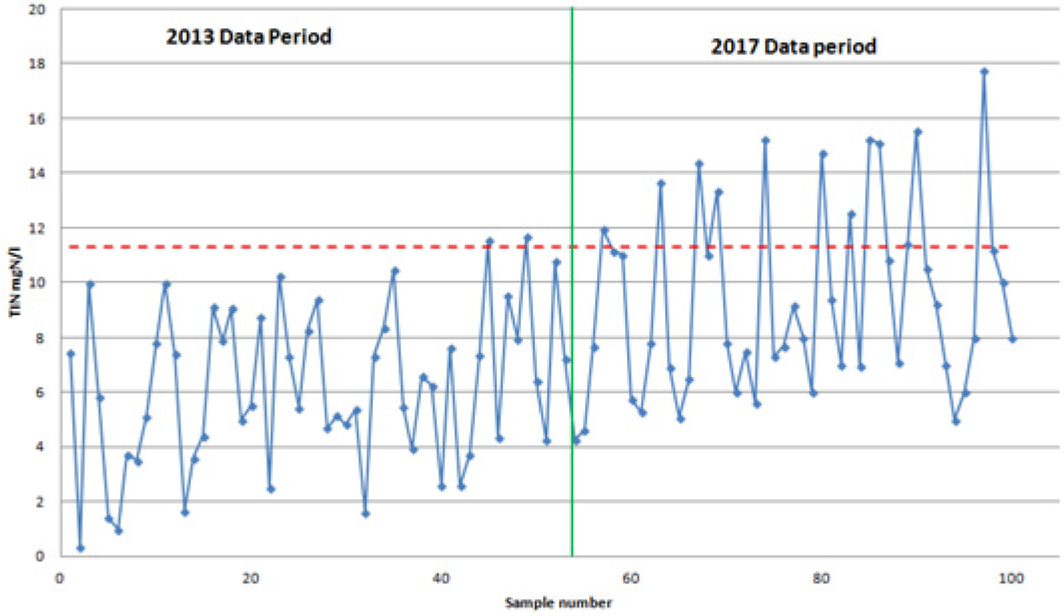


Figure 8.2 - Example time series TIN data for the case study monitoring point

More detailed analysis of the impact of the increasing load from the STW shows that the proportion of the N load it contributed increased from 30% to 67%. It is therefore now the dominant source of TIN in the WFD river catchment.

Despite this, the catchment would still be designated as there is clear evidence of deterioration at the site, from a marginal pass to a clear fail; the agricultural component is also increasing (in load terms if not proportionally) and contributes significantly to the TIN at this site.

9. Integrating areas of change into existing NVZ boundaries

Key concept 2.1 describes the review of NVZs as a continuous, rolling process. The majority of current NVZs were designated in 2005 or later, and subsequent reviews have concluded that the evidence still supports the view that the NVZ is needed. In these cases, the NVZ boundaries stay the same, except for minor changes due to appeals or changes in the field boundary dataset used to map the NVZ to actual land parcels.

There are five reasons why NVZ areas will change as a result of the current review:

1. we add a new NVZ because TIN levels are deteriorating
2. we remove an NVZ because it meets the de-designation criteria
3. we alter the area covered by an NVZ as we have better spatial data

4. a new, larger catchment designation replaces all other designations

5. a designation based on monitoring and modelling or monitoring only replaces an existing model only designation

The map below shows an example where we update existing water body designations with a new catchment NVZ. The yellow area represents the new NVZ while the hatched area beneath it is the existing (2013) NVZ. Despite the fact that the old NVZ covers the WFD river catchment at the outlet of the Rother, the previous failures are all situated on tributaries and so the previous designations were waterbody NVZ.

The evidence from the current review shows that the monitoring point at the catchment outlet, which has not failed in previous reviews, now exceeds the threshold TIN levels. Although the failing monitoring point is within an urban area, it meets the criteria for inclusion in this assessment and therefore we upgrade the historic water body designation to a new catchment NVZ as described in point 4 above.

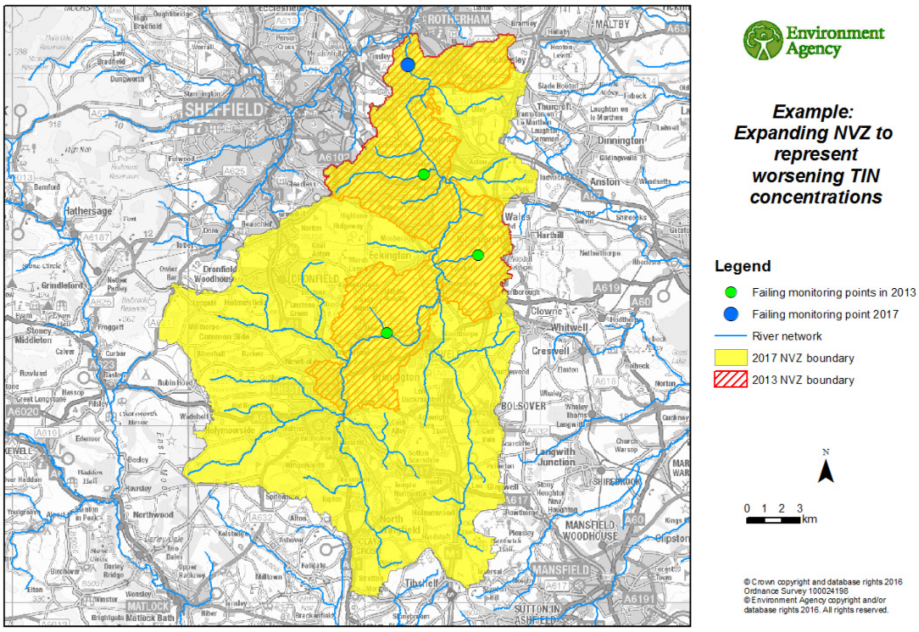


Figure 9.1 - Example of how we expand existing NVZ to take account of worsening TIN levels (River Rother)

9.1 Rules for removing designations

This is discussed in section 7 but if an area requires de-designation, we apply the following rules:

1. In a catchment NVZ, we will not remove parts of an NVZ upstream of the original designation if that downstream WFD river catchment does not meet the criteria for

de-designation, because these upstream areas still drain to water we consider at risk of being polluted.

2. When an existing NVZ is de-designated, any upstream areas that meet the criteria for **designation based on the current review** will be added as a new designation.

9.2 Modification of boundaries to better represent hydrology

The MRG agreed that for the 2009 and subsequent reviews the basic unit of assessment for surface water NVZs would be the WFD river catchment reporting units. The 2009 review was based upon the version of this dataset that predated Cycle 1 River Basin Management Plans (RBMPs) and was the version used for the initial River Basin Characterisation (RBC) processes. This split the surface area of England and Wales into 7,816 reporting units.

Since this review, these units have been reviewed and revised for two Cycles of RBMPs. In most places, the differences are minor but in certain areas (in particular the land draining to the Wash) the changes made in the latest Cycle 2 river catchment dataset are significant. They are, however, based on improved spatial data and a greater understanding of the hydrological processes in these areas and thus represent our best representation of the surface water catchments.

NVZ designations are a continuous process, but the majority of current designations are founded in the 2009 review and therefore based on the pre-RBMP dataset. In some places the external boundaries and connectivity between catchments will have been updated since the original NVZs were designated.

Updating all the existing boundaries to the new Cycle 2 equivalents is a complicated process involving areas of land not previously designated being included within NVZs, or excluded from an NVZ they had previously been allocated to. The 2009 review included a process to update all existing boundaries to the new preferred spatial units, so there is a precedent for such a large change. This decision, however, was discussed at length and agreed by the MRG of the time.

For the 2017 review we adopt a pragmatic approach to incorporating the new data. This approach means that:

- All new NVZs are based on Cycle 2 reporting units
- We update existing NVZs to Cycle 2 boundaries where the new boundaries represent a clear improvement over the old version.

In the example below, we see how the change to Cycle 2 boundaries results in a far more representative NVZ. Prior to this review this part of the Severn was a water body NVZ based on the continued failure of the monitoring point shown here as a green circle. The hatched area shows the extent of the historic NVZ. The revised Cycle 2 WFD river catchments distinguish between the land draining to the failing monitoring point and the land that drains only the main channel. We therefore remove the area that has no influence on the failing monitoring point and revise the NVZ to produce a more accurate representation of the catchment.

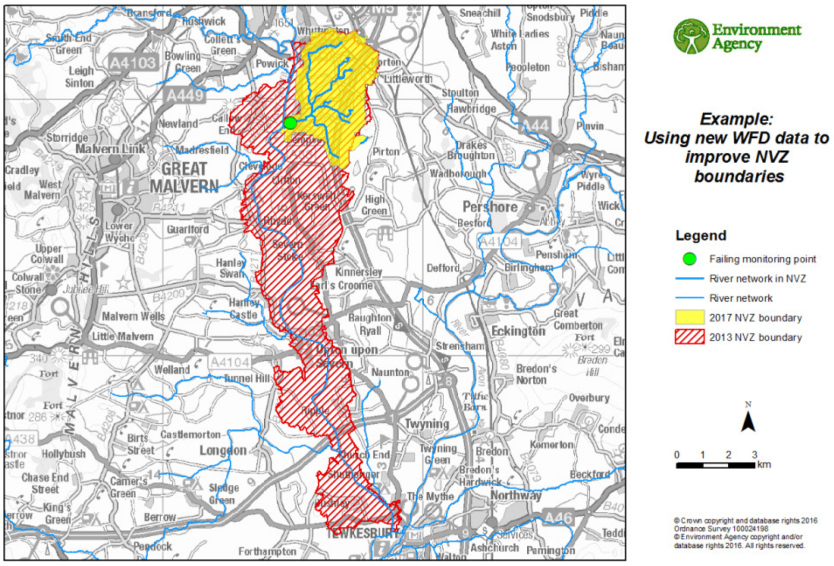


Figure 9.2 - Example of how using the revised WFD data can improve the fairness of NVZ boundaries (River Severn)

Another area where the update to revised WFD spatial data can result in more representative NVZs is in coastal catchments. The following example illustrates this. Again the hatched area represents the existing NVZ and the yellow area shows the revised boundary. The map in Figure 9.3 below clearly shows how the new boundary represents the land draining to the polluted water body without also designating large areas of unconnected land.

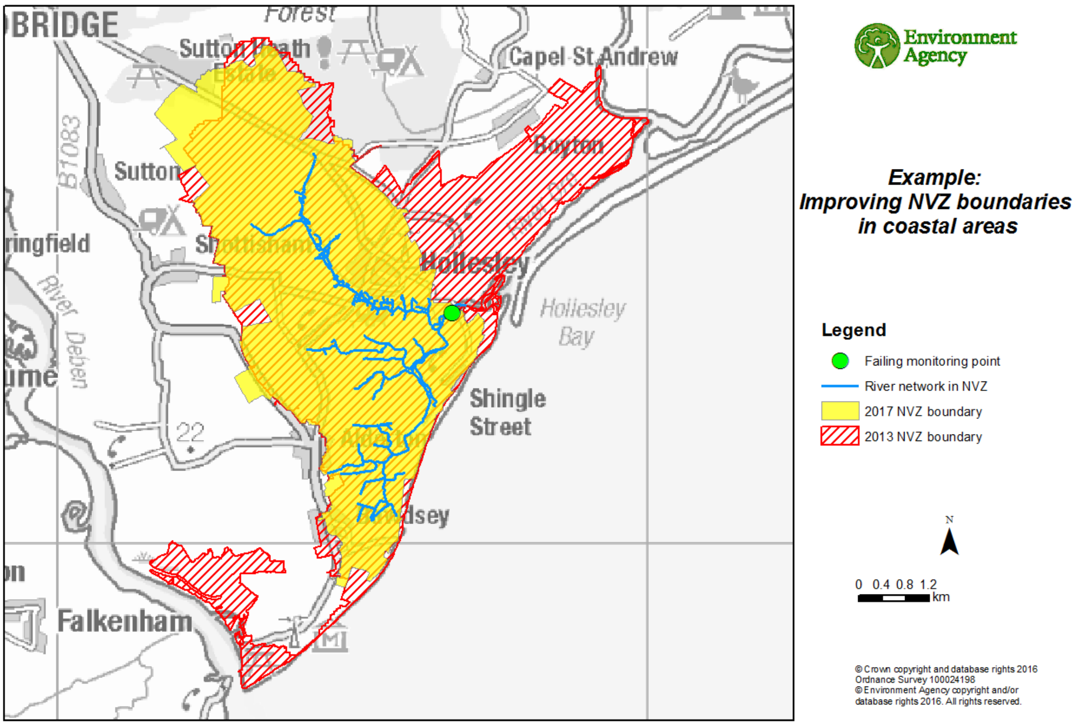


Figure 9.3 - Using revised WFD data to improve the accuracy of NVZ boundaries in coastal areas

The final type of water body that we are using updated WFD spatial data to better represent the land draining to the polluted waters is those managed by Internal Drainage Boards (IDBs) in the East of England.

For the 2013 review we acknowledged the difference between natural and operational drainage in areas of Lincolnshire, Cambridgeshire and Norfolk. The appeals of the 2009 review highlighted areas where the WFD river catchments do not represent the day-to-day flow of water. In 2013 we used a combination of WFD river catchments and Internal Drainage District spatial data to identify NVZs in the land draining to the Wash. This was not wholly successful as the drainage districts were also amalgamated management units and therefore made no real improvement to the Cycle 1 WFD river catchments.

As part of the update to the Cycle 2 RBMP reporting units, there was extensive consultation between the Environment Agency and local IDB operators in the areas we are most concerned with here (Lincolnshire, Cambridgeshire and Norfolk) and this has resulted in more representative WFD river catchments for rivers draining to the Wash.

For the 2017 review we adopt the following approach in areas draining to the Wash:

1. All existing NVZs are updated to Cycle 2 WFD river catchment boundaries

2. Any existing NVZ that the Cycle 2 WFD river catchment dataset identifies as draining only to a transitional/coastal water body (and therefore not a freshwater) is excluded, which are covered by the eutrophic NVZ method rather than the surface water NVZ method.

Almost all rivers draining to the Wash show strong evidence of pollution, so the fact that small parcels of land are not allocated to the right catchment is irrelevant if the adjacent area is also a NVZ. Using the updated data does allow us to identify any coastal or peripheral land that does not drain to polluted waters. Figure 9.4 below shows how this has resulted in a more logical NVZ in this area.

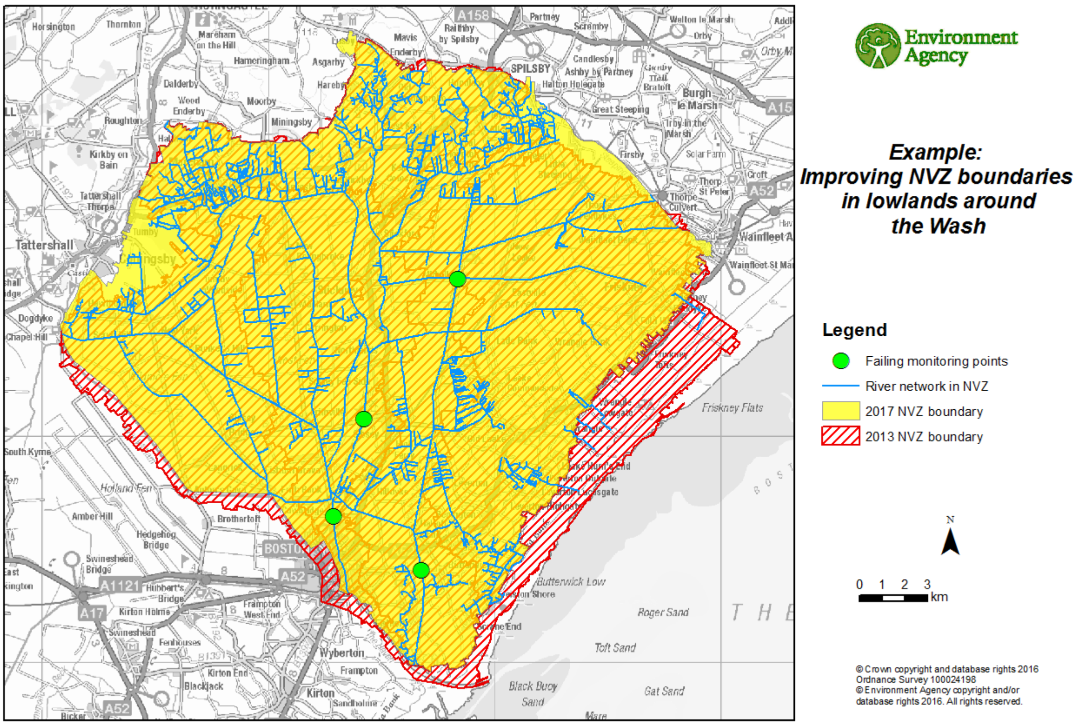


Figure 9.4 - Example of how we use new data to improve our representation of complex drainage for NVZ boundaries in lowland catchments draining to the Wash

We must note that the 2013 MRG supported the concept of a full shift from Cycle 1 WFD river catchments to the Cycle 2 equivalents, in advance of the data being available. Every effort has been made in the current review to implement this recommendation but fully implementing this recommendation has not been practicable. To do so would result in more changes to the coverage of NVZ that we believe the MRG anticipated when it made the decision. We therefore propose to show a future MRG the impact of a complete change from Cycle 1 to Cycle 2 catchments to inform a decision.

References

Alfons, Andreas. 2012. cvTools: Cross-Validation Tools for Regression Models. R Package Version 0.3.2.

Anthony, S., Quinn, P. and Lord, E. (1996) Catchment scale modelling of nitrate leaching. Asp. Appl. Biol., 46, 23-32.

Boorman, D.B., J.M. Hollis, and A. Lilly. 1995. Hydrology of Soil Types: A Hydrologically-Based Classification of the Soils of United Kingdom. http://nora.nerc.ac.uk/7369/.

Burgess, C.S. 2013. Developing a national dataset showing the likely location of septic tanks and other small sewage discharges. *Unpublished.*

Carey, Mike, and Paul Davidson. 2010. "Cumulative Nitrogen and Phosphorus Loadings to Groundwater." (November).

Chapra, S. Surface Water Quality Modelling 1997 McGraw-Hill ISBN 0-07-011364-5 0-07-843306-1

Doneker, R.L. and G.H. Jirka, "CORMIX User Manual: A Hydrodynamic Mixing Zone Model and Decision Support System for Pollutant Discharges into Surface Waters", EPA-823-K-07-001, Dec. 2007.
Available for download at HTTP://WWW.MIXZON.COM/DOWNLOADS/

Eatherall A, D.B. Boorman, R.J. Williams and R. Krowe 1998 Modelling in-stream Water Quality in LOIS Science of the Total Environment 210/211:499-517

EEA, 2000. CORINE:Co-ordination of information on the environment. Technical guide. European Environment Agency, Copenhagen.

Ellis, J.C., van Dijk, P.A.H. and Kinley, R.D. 1993. Codes of Practice for Data Handling – Version 1. WRc Report No. NR 3203/1/4224 for The National Rivers Authority.

Environment Agency 111_07_SD02 Water quality planning: Codes of practice for data handling

Environment Agency. 2012. Method Statement for Nitrate Vulnerable Zone Review – Surface Waters.

Environment Agency. 2016. NVZ Surface water land use modelling. Unpublished.

EU 2010 Common Implementation Strategy Guidance on setting mixing zones under the EQS Directive (2008/105/EC)

Griffiths, J, V Keller, D Morris, and A. R. Young. 2008. Environment Agency Science Report SC030240 Continuous Estimation of River Flows (CERF).

Holmes, M. G R, A. R. Young, T. H. Goodwin, and R. Grew. 2005. "A Catchment-Based Water Resource Decision-Support Tool for the United Kingdom." Environmental Modelling and Software 20: 197–202.

Koenker, R. and Hallock, K.F. 2001. Quantile regression. Journal of Economic Perspectives, 15 (4), 143-156.

Lee D., Gooday R., Steven A., Whiteley I. (2015), "Modelling Nitrate Concentrations in Groundwater: Neap-N Nitrate leaching for 1970 and 2014". Project ref: WT1550 ADAS report for Defra and the Environment Agency

Lord, E. and Anthony, S. (2000) MAGPIE: A modelling framework for evaluating nitrate losses at national and catchment scales. Soil Use and Management, 16, pp. 167-174.

Mitchell, G., Lockyer, J. And McDonald. A.T. 2001. Pollution hazard from urban non-point sources: A GIS-model to support strategic environmental planning in the U.K. Technical report, School of Geography, University of Leed, 122 pp.

Mitchell, G. 2005. Mapping hazard from urban non-point pollution: a screening model to support sustainable urban drainage planning. Journal of Environmental Management 74, 1-9.

Mitchell, G. 2006. Urban diffuse pollution: a basin scale perspective on screening and the role of urbanisation. Report to the Scottish Environment Protection Agency (SEPA)

Nitrate Pollution Prevention Regulations 2015. Statutory Instruments 2015 no.668.

Pygott, K. 1999. Alternatives to Dry Weather Flows in Setting Discharge Consents.

R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

USEPA Rates, Constants and Kinetics in surface Water Modelling (Second Edition) http://nepis.epa.gov/Adobe/PDF/9100R3IW.PDFWade, A. J., Butterfield, D. and Whitehead, P. (2006) Towards an improved understanding of the nitrate dynamics in lowland, permeable river-systems: Applications of INCA-N. Journal of Hydrology, 330 (1-2). pp. 185-203. ISSN 0022-1694 doi:10.1016/j.jhydrol.2006.04.023

WRc. 2015. Statistical methods for Nitrate Vulnerable Zone Review 2017. WRc report reference UC10943.07 *Unpublished.*

Appendix A. WIMS data extraction criteria

Data extraction criteria used in obtaining water quality data from the Environment Agency's Water Information Management System (WIMS) database.

Table A1 Determinand code - these describe the compound sampled

Determinand Code	Determinand Name
0111	Ammoniacal nitrogen as N
0116	Total oxidised nitrogen as N
0117	Nitrate as N
0118	Nitrite as N

Table A2 Purpose codes - these describe the reason the sample was collected

Purpose Code	Purpose Code Description
CA	Compliance audit (permit)
CI	Statutory audit (operator data)
CO	Water quality UWWTD monitoring data
CS	Water quality operator self-monitoring compliance data
IA	IPPC/IPC monitoring (Agency audit - permit)
II	IPPC/IPC monitoring (Agency investigation)
IO	IPPC/IPC monitoring (operator self-monitoring data)
MN	Monitoring (national Agency policy)
MP	Environmental monitoring (GQA & RE only)
MS	Environmental monitoring statutory (EU directives)
MU	Monitoring (UK govt policy - not GQA or RE)
PN	Planned investigation (national Agency policy)

Table A3 Sample point type - these describe the physical medium collected at a sample point and a broad typology

Point Type	Sampling Point Description
F1	Freshwater - RQO RE1
F2	Freshwater - RQO RE2
F3	Freshwater - RQO RE3

Point Type	Sampling Point Description
F4	Freshwater - RQO RE4
F5	Freshwater - RQO RE5

Appendix B. Water company monitoring data

Defra invited water supply companies to provide surface water quality data collected as part of their routine monitoring programmes via the industry body Water UK. Six companies provided valid data for a total of 108,721 samples from 251 unique surface water monitoring sites. Other water supply companies returned data that was not usable, mainly because it did not contain location information, or the location information was not accurate enough. Where returned data is expressed as nitrate, rather than as N, it was converted before being used in the analysis.

Any returned data that had been blended or mixed and therefore did not represent water quality from a single location was not used. Any data from water treated in a way that might affect the concentrations of nitrogenous compounds was also removed.

Table B1 - Summary of returned water company data

Water Company	Determinands recorded	Sites	Samples	Date span
Affinity Water	Nitrate as N, Nitrite as N, Nitrogen, Total Oxidised as N	4	1,770	Jan 2004 – June 2015
Bristol Water	Nitrate as NO_3	10	9,091	July 1995 – March 2015
Sutton and East Surrey	Nitrate as NO_3	1	219	Jan 2009 – Dec 2014
United Utilities	Ammoniacal Nitrogen as N Nitrate as N Nitrite as N Nitrogen, Total Oxidised as N	179	54,530	Jan 1980 – April 2015
Wessex Water	Nitrate as N	48	35,782	Jan 1980 – March 2015
Yorkshire Water	Nitrate as N	9	7,329	Jan 1980 – April 2015
Totals		**251**	**108,721**	

Appendix C. Multiple Outlier Test

C1 Objective

The objective of the Multiple Outlier Test (MOT) is to identify outliers in a given data set, on the assumption of underlying Normality.

C2 Definitions

Outlier: a data value which has arisen from some statistical population that is more extreme than the population from which the bulk of the values have arisen

Suspected outlier: a data value which is so far above or below the bulk of the data values that it causes surprise to the user of the data

Although the definition of suspected outlier might appear rather subjective, it carries with it the implication that *the user must have some correct probability distribution in mind* (however vague), and believes that the suspected outliers are not consistent with that distribution. In other words, he or she suspects that the sample has been contaminated by observations from some statistical distribution other than the one expected.

C3 The single outlier test

A well-established statistical procedure is available when the data can be assumed to have come from an underlying Normal distribution (or where the data can be transformed, for example by taking logarithms) so as to make this assumption reasonable. The test proceeds as follows. First calculate the mean (m) and standard deviation (s) of the data values. Then calculate the quantity t_{max} as:

$t_{max} = |(x? – m)|/s$

where x? is the suspected outlier (that is, either the minimum or the maximum of the data set).

If t_{max} is greater than the value given in Table C1, the outlier can be declared to be statistically significant at the 1% level. In other words, the probability that a value as extreme as this could have arisen by chance from a Normal population is only 1 in 100.

Table C1 - Critical values (P = 1%) of the tmax statistic

No. of data values	**Critical value (tmax)**
4	1.49
5	1.75
6	1.94
7	2.10
8	2.22
9	2.32
10	2.41
12	2.55
14	2.66
16	2.75
18	2.82
20	2.88
30	3.10
40	3.24
50	3.34
60	3.41
80	3.53
100	3.60
120	3.66
150	3.72
200	3.81
300	3.91
400	3.97
500	4.03

C4 The Multiple Outlier Test

What if several outliers are suspected to be present in the data? In that case a simple generalisation of the preceding test can be used. This is known as the Multiple Outlier Test.

The test takes an 'outward consecutive' approach. First, a pool of k suspects is produced by finding all the data values whose t_{max} value (as defined above) is greater than some arbitrary limit (see below). The k suspects are then tested one at a time, working from the least to the most extreme. The details of the procedure are as follows:

- Starting with the (n-k) 'reliable' data values, augment these by just the least extreme of the k suspected outliers.
- Calculate the mean and standard deviation of those (n-k+1) values, and perform the single outlier test as usual.
- If the suspect fails to be confirmed as an outlier, pool it with the (n-k) reliable values, recalculate the mean and standard deviation, and then test the next least extreme suspect.
- Continue in this way until a suspected outlier is declared to be a genuine outlier. All the remaining (and hence more extreme) values are then declared to be outliers also.

It might be thought that, as the multiple outlier test provides several opportunities for false positives, the actual significance levels would be somewhat higher than the nominal 1% value quoted in Table C1 for the single outlier case. That is not the case, however. When the data values really do come from a Normal population, the multiple outlier test very rarely produces false positives unless the single outlier test does so also - a characteristic that has been confirmed by computer simulation.

Following the 2008 and 2013 Review methodologies, the nitrate datasets were assumed to be Normally distributed. A critical value of 7 was used to strike a reasonable balance between identifying outliers that were plainly abnormal and retaining data values that were merely suspicious.

Appendix D: Note on the Weibull and Quantile regression methods

D1 The Weibull method

The Weibull method uses the r^{th} ranked value within the observation dataset to provide an estimate of the 95^{th} percentile, where r = 0.95(n + 1) and n is the number of samples. When r is not an integer, r is rounded down and up to the nearest whole number, and the corresponding concentration values for these ranks are interpolated to estimate the 95^{th} percentile. Conservative 90% and 50% confidence intervals are calculated using binomial distribution theory, as described in the Environment Agency Codes of Practice for Data Handling (Ellis et al. 1993). A minimum of 28 and 59 samples are required to calculate the upper 50% and 90% confidence limits, respectively, so for some sites it was possible only to demonstrate with medium or low confidence that the 95^{th} percentile was below the threshold. If the lower 90% confidence limit exceeds 11.3 mgN/l TIN as N, the monitoring point is deemed to have failed the test with high confidence; if the lower 50% confidence limit exceeds 11.3 mgN/l TIN as N, the monitoring point is deemed to have failed the test with medium confidence; if the 95^{th} percentile estimate exceeds 11.3 mgN/l TIN as N, the monitoring point is deemed to have failed the test with low confidence.

D2 Quantile regression

Quantile regression (Koenker and Hallock 2001) is a statistical technique that explores how one or more independent variables influence a specified percentile value of the response variable, for example the median (the 50^{th} percentile). In contrast to conventional linear regression, which seeks to explain variation in the mean of the response variable, quantile regression can explain variation in percentile values of the response variable (95^{th} percentile TIN as N concentrations in this case).

Quantile regression is a robust technique that makes no assumptions about the underlying distribution of the data. It is also relatively insensitive to outliers.